Expression of gender in the human voice:
investigating the “gender code”

Dissertation submitted to the University of Sussex for the degree of Doctor of Philosophy

Valentina Cartei

April 2014

UNIVERSITY OF SUSSEX

Valentina Cartei **Doctor of Philosophy**

The expression of gender in the human voice: investigating the "gender code"

Summary

We can easily and reliably identify the gender of an unfamiliar interlocutor over the telephone. This is because our voice is "sexually dimorphic": men typically speak with a lower fundamental frequency (F0 - lower pitch) and lower vocal tract resonances (ΔF – "deeper" timbre) than women. While the biological bases of these differences are well understood, and mostly down to size differences between men and women, very little is known about the extent to which we can play with these differences to accentuate or de-emphasise our perceived gender, masculinity and femininity in a range of social roles and contexts.

The general aim of this thesis is to investigate the behavioural basis of gender expression in the human voice in both children and adults. More specifically, I hypothesise that, on top of the biologically determined sexual dimorphism, humans use a "gender code" consisting of vocal gestures (global F0 and ΔF adjustments) aimed at altering the gender attributes conveyed by their voice. In order to test this hypothesis, I first explore how acoustic variation of sexually dimorphic acoustic cues (F0 and ΔF) relates to physiological differences in pre-pubertal speakers (vocal tract length) and adult speakers (body height and salivary testosterone levels), and show that voice gender variation cannot be solely explained by static, biologically determined differences in vocal apparatus and body size of speakers. Subsequently, I show that both children and adult speakers can spontaneously modify their voice gender by lowering (raising) F0 and ΔF to masculinise (feminise) their voice, a key ability for the hypothesised control of voice gender. Finally, I investigate the interplay between voice gender expression and social context in relation to cultural stereotypes. I report that listeners spontaneously integrate stereotypical information in the auditory and visual domain to make stereotypical judgments about children's gender and that adult actors manipulate their gender expression in line with stereotypical gendered notions of homosexuality. Overall, this corpus of data supports the existence of a "gender code" in

human nonverbal vocal communication. This “gender code” provides not only a methodological framework with which to empirically investigate variation in voice gender and its role in expressing gender identity, but also a unifying theoretical structure to understand the origins of such variation from both evolutionary and social perspectives.

Acknowledgements

Firstly, I extend my sincere gratitude and appreciation to my principal supervisor David Reby, for his expert guidance, practical support and unwavering encouragement throughout this research programme. Many thanks also to the other co-supervisors, Robin Banerjee, Karen McComb, and Jessica Horst for their constructive feedback and their technical assistance on experimental design, participant recruitment and statistics. I thank Pennie Ingram for her administrative support, and Jennifer Rusted for her guidance and support in getting Ethics approval for the research. I would also like to thank my lab colleagues for their assistance, intellectual stimulation and fellowship. I extend a special thanks to a fellow office colleague and friend, Victoria Ratcliffe, who for two years has offered continuous support on both a scholarly and personal level. Finally, infinite thanks to my cheering squad: my family, friends and work colleagues at Survivors' Network. I thank them deeply for their emotional support and understanding.

Declaration

The thesis conforms to an “article format” in which the middle chapters consist of discrete articles written in a style that is appropriate for publication in peer-reviewed journals in the field. The first and final chapters present synthetic overviews and discussions of the field and the research undertaken.

Study 2 in Chapter 3 is published in *PLoS ONE* as: “Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children’s Voices”

The authors’ contributions are as follows: I was responsible for the study design, data collection, data analysis and writing of the manuscript; Dr. Reby and I were responsible for initial conception of the research; Dr. Reby contributed to the study design and provided feedback on the manuscript.

Study 3 in Chapter 3 is under revision for the *Journal of Hormones and Behavior* as: “What makes a voice masculine: multilevel investigation of physiological and acoustical bases of perceived masculinity”

The authors’ contributions are as follows: I was responsible for all aspects of the study from its initial conception to writing of the manuscript; Dr. Bond provided feedback on study design and data analysis; Dr. Reby contributed to the study design, data analysis and provided feedback on the manuscript.

Study 4 in Chapter 4 is published in the *British Journal of Developmental Psychology* as: “Control of voice gender in pre-pubertal children”

The authors’ contributions are as follows: I was responsible for most aspects of the study, including study design, data collection, data analysis and writing of the manuscript; Dr Reby and Dr. Cowles share responsibility for the initial conception of the study. Dr. Reby and Dr. Banerjee provided feedback on study analysis and corrections to the manuscript.

Study 5 in Chapter 4 is published in *PLoS ONE* as: “Spontaneous Voice Gender Imitation Abilities in Adult Speakers”

The authors' contributions are as follows: I was responsible for most aspects of the study, including study design, data collection, data analysis and writing of the manuscript; Dr. Cowles and Dr Reby share responsibility for the initial conception of the study. Dr. Reby contributed to the study design, data analysis and provided feedback on the manuscript.

Study 8 in Chapter 5 is published in the *Journal of Nonverbal behavior* as: "Acting Gay: Male Actors Shift the Frequency Components of Their Voices Towards Female Values When Playing Homosexual Characters"

The authors' contributions are as follows: I was responsible for all aspects of the study from data collection to writing of the manuscript; Dr Reby and I were responsible for the initial conception of the study. Dr. Reby contributed to the study design, data analysis and provided feedback on the manuscript.

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:………………………………………

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Overview

Research on human vocal communication traditionally focuses on speech and its role in communicating linguistic information (Hewlett & Beck, 2013). More recently, however, the nonverbal dimension of speech signals has received growing attention, with studies highlighting how the voice can provide cues to many different dimensions of speakers including their emotions (Yogo, Ando, Hashi, Tsutui & Yamada, 2000); personality traits (Aronovitch, 1976; Scherer, 1979); attractiveness (Berry, 1992; Collins, 2000); maturity (Berry, 1992; Hummert, Mazloff & Henry, 1999; Mulac & Giles, 1996); age (Bruckert, Liénard, Lacroix, Kreutzer & Leboucher, 2006; Collins & Missing, 2003) and occupation (Yamada, Hakoda, Yuda & Kusuhara, 2000).

One of the key characteristics of the human voice is the existence of marked differences between men’s and women's voices. While there is some evidence that men speak with a less breathy, more creaky, and more monotonous voice than women’s (Henton, 1995; Klatt & Klatt, 1990; Mendoza, 1996), the most well-documented sex differences are in voice fundamental frequency (associated with the percept of pitch) and overall spacing of vocal tract resonances or formants (ΔF – associated with the percept of timbre): men have disproportionately lower-pitched and more resonant (deeper) voices than women (Titze, 1994). Besides differences in body size (men are 20% heavier (Hollien, 1960) and 7% taller than women (Gaulin & Boster, 1985)), this acoustic dimorphism in F0 and ΔF is largely based on men developing an enlarged larynx (producing lower F0) and an elongated vocal tract (producing lower ΔF) during puberty (Titze, 1994). Acoustic variation in F0 and ΔF between the two sexes as well as between individuals of the same sex, suggests that in addition to this biologically based variation, a proportion of sex differences may be behavioural in origin (Johnson, 2006; Sachs, Lieberman & Erickson, 1973; Whiteside, 2001).

There has been a recent surge in voice-related research investigating the anatomical and behavioural origins of between and within-sex differences in voice F0 and ΔF from an evolutionary perspective: e.g. showing that lower-pitched, more resonant voices give males a competitive advantage in intimidating rivals and/or attracting mates (Hodges-Simeon, Gaulin & Puts, 2011). Acoustic signals are sexually dimorphic in many species (for a review see Andersson, 1994), with sexually mature males having disproportionally larger vocal apparatuses and thus producing lower frequency calls than females (baboons: Rendall, Kollias, Ney & Lloyd, 2005; red deer: Fitch & Reby, 2001; fallow deer: McElligott, Birrer & Vannoni, 2006; Mongolian gazelle: Frey, Volodin, Volodina, Soldatova, & Juldaschev, 2008; Old World and Asiatic leaf monkeys: Dixson, 2012). Sexually-selected voice components (fundamental frequency and resonance frequencies) can also cue to key ecological traits: e.g. lower-pitched, more resonant vocalisations are typically associated with larger and/or higher quality males (North American bison: Wyman et al., 2011; red deer: Reby & McComb, 2003; rhesus macaques: Fitch, 1997; giant pandas: Charlton, Zhihe, & Snyder, 2009; Charlton et al., 2011) and are considerably more effective in deterring rivals and/or attracting mates (red deer: Reby et al., 2005; Charlton, Reby & McComb, 2007; domestic dog: Taylor, Reby & McComb, 2010; Australian sea lions: Charrier, Ahonen, & Harcourt, 2010; giant pandas: Charlton et al., 2009). Besides the anatomical adaptations underlying the observed acoustic variation in F0 and ΔF, studies have shown that callers have evolved behavioural strategies that enable them to alter the relationship between biological attributes and voice frequencies of their vocal signals in order to influence the outcome of mating and/or competitive contexts. In particular, several studies in recent years have provided support for the “size code” hypothesis (Ohala, 1984), showing that human and non-human males exploit the relationship between voice frequencies and body size by dynamically changing their frequencies in order to influence attributions of size and associated traits, e.g. lowering their formant spacing and F0 to sound bigger, more dominant and/or more aggressive (ΔF: red deer: Reby & McComb, 2003; fallow deer: McElligott et al., 2006; Mongolian gazelle: Frey et al., 2008; humans: Puts, Gaulin & Verdolini, 2006; wapiti: Fitch & Reby, 2001; F0: male white-lipped frogs: Lopez, Narins, Lewis & Moore, 1988).

However, unlike most mammalian mating calls, the human voice is used in a wide range of social contexts and cannot be reduced to a mating signal. Throughout this thesis, I will argue that sexually selected voice components (fundamental frequency and resonance frequencies) do not only express biological traits which characterise an individual as male or female (e.g. one’s *sex*), but also the socially and culturally constructed meanings that a given society, in a given time frame, associates with being, and behaving as, a man or a woman – commonly termed *gender* (Money, 1955). Further, I will show that speakers manipulate sexually dimorphic voice cues in order to vary the perceived gender of their voice and related attributes. Indeed, research in non-auditory domains shows that individuals of both sexes can exhibit feminine (e.g. gentleness, dependence, sensitivity) and masculine (e.g. dominance, self-reliance) traits and that they can voluntarily vary the expression of these traits, for example by changing mannerisms, clothing, and hairstyle (Maltry & Tucker, 2008; Nanda, 1999). With regards to the voice, this means that, for instance, while women have overall higher voices than men, a woman who speaks with a low voice may also consciously or unconsciously project a different gender image than a woman who speaks with a high voice. In this regard, the most cited and notorious example is perhaps Margaret Thatcher's use of speech therapy to lower her voice in order to sound more authoritative and masculine (Graddol & Swann, 1989). Yet, while isolated examples provide us with anecdotal evidence for a role of vocal behaviours in the context of gender expression, this area has received – surprisingly – very little scientific attention, and both its nature and role in human speech remain to be systematically investigated.

The central argument of this thesis is that speakers vary their sexually dimorphic acoustic cues (F0 and ΔF) along their biologically based polarity, in order to vary the expression of their gender through the voice (e.g. their maleness, femaleness, masculinity, femininity). The methodological background for this research is the source-filter theory of voice production (Fant, 1960), presented at the beginning of this Introduction. By decomposing the acoustic structure of vocal signals according to their mode of production, the source-filter theory provides a unifying framework to understand how acoustic variation is linked to (and likely to encode) anatomical or biological attributes of the caller (Taylor & Reby, 2010). Building on the basic understanding of how the voice is produced from a source-filter perspective, the present introduction offers an overview of how anatomical sex differences relate to acoustic sex differences across the individual's lifespan, highlighting that acoustic variation cannot be fully explained by biological factors. Based on this observation and on the "size code" hypothesis also further detailed in this chapter, I review preliminary evidence for the existence of a "gender code", by looking at humans' ability to vocally imitate the opposite gender (e.g. in acting contexts and in real life), and at sociocultural differences in voice gender expression which exceed differences in size, while reflecting differences in gender roles (e.g. “gay” speech styles, society-specific definitions of “masculinity” or “femininity”). A summary of the research questions and thesis outline conclude the present chapter.

The Source-Filter Theory of Speech Production

According to the “source-filter theory” framework (Fant, 1960), the production of voiced signals in human speech follows a two-stage process: firstly, a signal is generated by a “source” and then passes through a “filter” which causes the signal to be modulated before being radiated out. The “source” is located at the level of the glottis (the vocal folds and the opening between them), where the signal is produced by periodic vibration of the vocal folds due to the continuous energy provided by the airflow passing through the glottis. This periodic oscillation creates a complex periodic wave whose spectrum contains a fundamental frequency, or F0, (equal to the rate of glottal vibration), and its integer multiple frequencies, the harmonics. Because vocal fold oscillation can be approximated by the behaviour of a simple vibrating string (Titze, 1989), F0 can be predicted by the following formula (1):

$$F0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}$$

where L is the vocal fold length, σ is the stress applied to the vocal fold (force per unit area), and ρ is the tissue density (1.02 g/cm^{-3}). Thus, F0 is inversely proportional to vocal fold length and directly proportional to the square root of tension on the vocal folds, with longer, heavier, and looser vocal folds vibrating at a lower fundamental frequency. The rate at which the vocal folds open and close during phonation can be varied in a number of ways and can be dynamically changed by the tension of the laryngeal muscles (mainly posterior cricoarytenoids – vocal fold abductors – and interarytenoids – vocal fold adductors) and the air pressure generated by the lungs. Perceptually, the fundamental frequency is responsible for the perceived “pitch” of the voice.

In the second stage, as the glottal wave propagates through the supra-laryngeal vocal tract (from the larynx to the lips), selected frequencies from the signal are dampened or amplified, producing spectral peaks called “formants” (Fi). Formants are mainly responsible for the perceived "timbre" of the voice and modulation of the first two formants are the main determinants of the different sounds that we perceive as vowels (Fant, 1960; Titze, 1994).

As a first approximation (Titze, 1994), the vocal tract can be modelled as an open tube closed at one end (the glottis) and open at the other (the mouth). Under this model, formant frequencies can be estimated by the following formula (2):

$$F_i = \frac{(2i-1)c}{4VTL}$$

where *c* is the speed of sound in air (approximated as 350m/s in the human vocal tract) and *VTL* is the length of the vocal tract. The above formula indicates that the primary determinant of Fi is the length of the vocal tract, with longer vocal tracts producing lower and more closely spaced formants, suggesting in turn that individual formants would provide an acoustic estimate of vocal tract length during phonation. In reality, the estimation of vocal tract length from individual formant frequencies is only accurate if the cross-sectional area of the tract is uniform, as in the "schwa" vowel. For all the other vowels the configuration of the vocal tract is more complex, and vocal tract size and shape, as well as length, affect formant frequency values (Fitch & Hauser, 2003). For example, while the vocal tract can be lengthened by lowering the larynx or protruding one's lips (thus lowering all formants), and shortened by raising the larynx or spreading one's lips (thus raising all formants, (Titze, 1994)), individual formant values are diversely affected by the place of constriction of the tongue body and tip in the oral and pharyngeal cavities (which changes the shape of the tract), and by the opening and closing of the mouth (which change the size of such cavities (Titze, 1994)). Individual formants are also affected by glottal state (e.g. a tube open at both ends, with the glottis not entirely closed, will have higher F1 than one with one end closed (Fitch & Hauser, 2003)). Formant spacing, the average distance (measured in Hertz) between successive formants, provides a better estimate of anatomical vocal tract length than individual Fi, as it is not affected by boundary (end) conditions (Fitch & Hauser, 2003). ΔF can be calculated as (3):

$$\Delta F = \frac{\sum_{i=1}^{N-1} F_{i+1} - F_i}{N-1}$$

where ΔF is the formant spacing (in Hz), Fi is the frequency of the ith formant and N is the total number of formants measured (adapted from Fitch, 1997). An alternative method, which will be used throughout this thesis, is the regression method of Reby and McComb (2003) in which ΔF is deduced from the equation for the quarter-wave length resonator described above (2), by plotting the observed frequency values against those that would be expected if the vocal tract was a straight uniform tube (further details in Chapter 2: "Materials and Methods"). This method describes ΔF in terms of its acoustic correlate, apparent Vocal Tract Length (aVTL), which is measured in cm, rather than in Hz, and thus provides an estimate of "speaking" VTL, the anatomical vocal tract length achieved during phonation, as opposed to "resting" VTL, which is the anatomical VTL achieved during quiet breathing. For the purpose of this thesis, ΔF will often be described in terms of aVTL as this estimate allows us to relate global formant shifts to the behavioural gestures (vocal tract lengthening or shortening) underpinning such shifts.

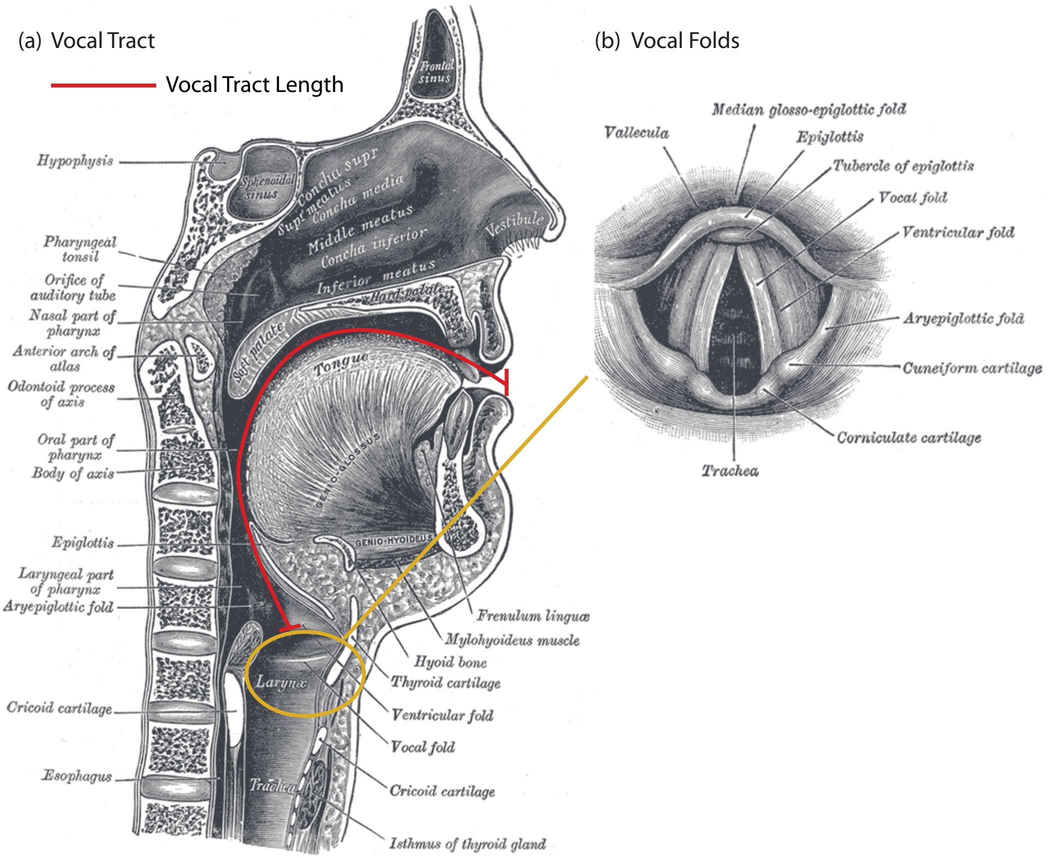

Figure 1.1 Sagittal view of the human vocal tract (a) and transverse view of the vocal folds (b), which are located within the larynx. The red line illustrates vocal tract length. Adapted from Gray's Anatomy of the Human Body (p.1079), by H. Gray, 1918, Philadelphia: Lea & Febiger. Copyright 1918 by Lea & Febiger.

Origins of the Sexually Dimorphic Voice

Sex Dimorphism of the Vocal Apparatus

From infancy to the onset of puberty, vocal folds and tract lengthen linearly with age in both sexes, despite localised sex differences in growth rate of vocal fold tissue (0.7mm in boys and 0.4mm in girls (Titze, 1994)), and growth rate and type of selected vocal tract sections (Vorperian et al., 2009; Vorperian et al., 2011). During puberty, however, androgen-related changes affect the male and female vocal apparatus differentially (Titze, 1994). More specifically, a surge in male androgen levels during this period causes a permanent enlargement of the male larynx (resulting in the noticeable protrusion of the thyroid notch, or Adam's apple) and a related 63% increase in the lengthening of the membranous portion of male vocal folds, whereas female vocal fold length increases by only 34% over the same period (Kahane, 1982). By the end of male puberty, men's vocal folds are therefore twice as long as females' vocal folds, lengthening from 4–8 mm at birth, to 29mm in adult males and 21mm in adult females (Kent and Vorperian, 1995; Linders, Massa, Boersma & Dejonckere, 1995). The pubertal increase in circulating levels of androgens also appears to underpin men's differential body height (men's bodies grow 7% more in height than women on average (Gaulin & Boster, 1985)) and a male-specific second large descent of the larynx occurring at puberty (Fitch & Giedd, 1999; Vorperian et al., 2009). These changes result in men developing vocal tracts that are 20% longer than women's on average (Fant, 1960), lengthening from about 8 cm in infancy, to 18 cm in adult males and 15cm in adult females (Vorperian et al., 2009).

Acoustic implications. In line with the source-filter theory, the aforementioned age- and sex- specific differences in the growth of the vocal apparatus have implications for its acoustic properties. At the level of the source, fundamental frequency declines during the course of development, consistent with concomitant increases in body size growth (Titze, 1994). Moreover, sex differences in F0 closely track concomitant sex differences in vocal fold length (see Table 1.1 for a summary of studies examining sex differences in mean fundamental frequency across the lifespan). It is generally reported that pre-pubertal boys and girls speak with the same F0 (Lee, Potamianos & Narayanan, 1999), reflecting the absence of significant differences in vocal fold length between the two sexes during that period. Sex differences in F0 start to appear around age 12, corresponding to the end of female puberty and the beginning of male puberty, when the male larynx starts to grow faster than the female larynx (Lee et al., 1999). By this age, females' F0 stops declining, reaching adult values of about 200 Hz, while males' F0 drops rapidly, reaching adult values of about 100 Hz by age 15 (end of male puberty (Lieberman, 1988)). As a result, post-pubertal males speak with a 50–80% lower F0 than females (Hollien, Green & Massey, 1994; Lee et al., 1999), a difference that remains unchanged throughout most of adulthood, in line with the absence of subsequent increases in vocal fold length for either sex (Lee et al., 1999; Titze, 1994). The F0 of males and females do converge again, however, from about 50 years of age, though this fact has been traced back to changes in vocal fold tissue rather than its length. More specifically, the reported 35Hz rise in men's F0 from middle age onwards (Hollien & Shipp, 1972; Krook, 1988; Van Rie & Van Bezooijen, 1995) has been attributed to ageing vocal folds thinning and deteriorating (Calhoun & Eibling, 2013; Deliyski, 2001; Linville, 2004). Similarly, the 10Hz drop in women's F0 after menopause (Honjo & Issiki, 1980; Russell, Penny & Pemberton, 1995; Torre III & Barlow, 2009) is consistent with a drop in their oestrogen and progesterone levels (Abitbol, Abitbol & Abitbol, 1999).

Table 1.1

A selection of studies (after 1990) that measured mean F0 (in Hz) on speakers across the lifespan

Age	Language	Sample size	Speech material	Male F0	Female F0	Difference	Author(s)	Year
0-9 months				u	u	ns	Kent	2002
3-8 months				300	300		Kuhl et al.	1996
3-4		10		257	249	ns	Trollinger	2003
5-6		10		257	243	ns		
7-8		10		234	253	ns		
5-7		25,23 a	vowel	240b		ns	Baker et al	2008
			phrase	237b		ns		
			sentence	236b		ns		
			counting 1-10	247b		ns		
5-11	American English	20-50c	vowels	260	268	ns	Lee et al	1999
12				226	231	∨		
15				127	226	∨		
16-18				127	228	∨		
24-25	Canadian English	20	spont /read speech	116	199	∨	Britto & Doyle	1990
20-35	American English	15	read	118	192	∨	Brown et al	1999
40-55		20		100	195	∨		
70-80	English	21, 23 [a]	vowel /a/	128	188	∨	Deliyski	2001

Note. The direction of a significant effect with respect to F0 is indicated with arrows, where ∨ shows that males have significantly lower F0 than females, and ∧, that males have significantly higher F0 than females. Peri- and post-pubertal males have significantly lower F0 than females, while no significant differences in F0 are reported before puberty. a. for unbalanced samples, the number of male and female participants is reported separately e.g. 25, 23 means 25 males and 23 females, b. values reported as average across genders, c. 20-50 children per year. Sample size varied according to vowel uttered and age group

The “source-filter” theory also predicts that a lengthening of the vocal tract leads to an overall decrease in its resonant frequencies and a narrowing of their spacing. Indeed, overall, vocal tract length scales with age-related body size growth in both males and females. Additionally, the documented growth spurt in vocal tract length observed during male puberty provides strong biological support to the faster and greater decrease in formant values of pubertal males relative to females (see Table 1.2 for a summary of studies examining sex differences in the first four formant frequencies across the lifespan). By the end of puberty, the male vocal tract is approximately 40% longer than children’s (Sudenberg, 1987, p.102 cited in Welch and Howard, 2002) and 20% longer than adult females’ (thus male voice formant spacing is about 80% of female's (Hillenbrand & Clark, 2009; Rendall et al., 2005)). It should be noted, however, that males' individual formant frequencies are not related to females' by a single downscaling scaling factor as sex differences also involve subtle differences in formant position reflecting sex variation in vocal tract morphology or gesture (discussed in more detail in Study 1). Moreover, in contrast with anatomical data (Fitch & Giedd, 1999; Vorperian et al., 2011), the sexual dimorphism in formant frequencies emerges long before the documented pubertal dimorphism in overall vocal tract length, with acoustic studies reporting lower (6-9%) values in boys' individual formants compared to girls' (Bennett, 1981; Bennett & Weinberg, 1979; Busby & Plant, 1995; Eguchi & Hirsh, 1968; Hasek, Singh, & Murry, 1980; Lee et al., 1999). While localised sex differences in vocal tract growth, rate and volume may also contribute to these differences (Vorperian & Kent, 2007; Vorperian et al., 2009; Vorperian et al., 2011), the anatomical origins of the pre-pubertal dimorphism in formant frequencies remain largely unknown. The unexplained mismatch between anatomical (resting) vocal tract length and its acoustic correlates throughout development has led several authors to suggest that acoustic variation in formant frequencies may also have a gestural, behavioural origin (Lee et al., 1999; Sachs et al, 1973; Whiteside, 2001).

Table 1.2

A selection of studies (after 1981) that measured formants (F1-F4 in Hz) on speakers across the lifespan

Age	Language	Sample size	Speech Material	Male vs Female differences				Author (s)	Year
				F1	F2	F3	F4		
4	American English	10	vowel /a/	ns	ns	ns		Huber et al.	1999
8		10		ns	∨	ns			
10		10		ns	ns	ns			
12		10		ns	ns	ns			
14		10		∨	∨	∨			
4	American English	10	vowels	ns	ns	∨		Perry et al.	2001
8-16		10a		∨	∨	∨			
5	Australian English	10b	vowels	∨	∧	∨		Busby & Plant	1995
7-11				∨	∨	∨			
5	English, Korean	20	vowels	∨	∨			Lee & Iverson	2009
10		20		∨	∨				
11+	American English	20	vowels	∨	∨	∨		Lee et al.	1999
6-10	American English	8c	vowels		∨	∨		Whiteside & Hodgson	1999
7-8	American English	42	vowels	∨	∨	∨	∨	Bennett	1981
18-44	Canadian English	34	vowels	∨	∨	∨	∨	Rendall et al.	2005

Note. The direction of a significant effect with respect to Fi is indicated with arrows, where ∨ shows males have significantly lower Fi than females, and ∧, that males have significantly higher Fi than females. Sexual dimorphism in Fi emerges by 4 years of age, with differences becoming more apparent by age 7, at which age boys have consistently lower formant frequencies than girls. a. 10 participants in each age group (8-, 12- and 16-year-olds). b. 10 participants in each age group (5-, 7-,9- and 11-year-olds). c. 8 participants in each age group (6-, 8- and 10-year-olds).

Perceptual implications. In view of the anatomical and acoustical dimorphisms previously described, researchers have focused on the extent to which listeners are able to identify the sex of speakers from their voices, and their use of F0 and ΔF cues when making sex attributions. It appears that listeners are able to discriminate the sex of their interlocutors from a very early age. Cross-modal studies report that four-month-olds look significantly more to the adult male face when the male voice is played and to the adult female face when the female voice is played (Walker-Andrews & Lennon, 1991), and that by seven months infants match faces and voices of nine-year-old children on the basis of the speaker's sex (Bahrick, Netto, & Hernandez-Keif, 1998). By the age of six, children's performance on voice sex identification of adult voices reaches the same high level of accuracy displayed by adult listeners (75% to 98% (Bennett & Montero-Diaz, 1982; Coleman, 1976; Hillenbrand & Clark, 2009; Hollien et al., 1994; Lass, Hughes, Bowyer, Waters & Bourne, 1976; Whiteside, 1998)). Moreover, while no study to date has investigated pre-pubertal children's ability to identify sex of pre-pubertal speakers, adult listeners have been found to be able to discriminate sex of unseen child speakers as young as four (Perry, Ohde & Ashmead, 2001; Weinberg & Bennett, 2005).

The perception of speaker sex closely follows the acoustic (and anatomical) dimorphism. Adult males are perceived to speak with lower-pitched (lower F0) and lower-resonance (lower formant values and narrower spacing) voices than their female peers (Hillenbrand & Clark, 2009) and sex identification ratings of adult voices are almost exclusively accounted for by the combined effect of those acoustic parameters (98.8% (Bachorowski & Owren; 1999; Coleman, 1976; Hillenbrand & Clark, 2009; Gelfer & Mikos, 2005; Pisanski & Rendall, 2011; Skuk & Schweinberger, 2013; Smith & Patterson, 2005)). Most studies also report that F0 plays the greatest role in cueing for the sex of adult speakers (Gelfer & Mikos, 2005; Hillebrand & Clark, 2009; Lass et al., 1976; Whiteside, 1998; but see Smith & Patterson, 2005), in line with the greater dimorphism of F0 over ΔF in adult voices (F0 ratio: 1.81 versus ΔF ratio: 1.20 (Hillenbrand & Clark, 2009; Titze, 1994)). Intriguingly, adult male voices are more readily identified than female voices. For example, Bennett & Montero-Diaz (1982) found that children accurately identified the speaker's sex as male in 93.6% cases against 84% when the speaker was a woman. Similarly, Owren, Berkowitz and Bachorowski (2007) reported that adult listeners attribute the sex of speakers more quickly and more accurately (96.5%) from male vowels than from female vowels (90.5%), relating this bias to the fact that the vocal apparatus of pubertal males diverges from a previously shared developmental trajectory with females. Compared to research on adult voices, sex perception in pre-pubertal voices remains a largely unexplored area. While correlational studies suggest that adult listeners use sex differences in vowel formant frequencies to identify children's sex from their voices (Perry et al., 2001; Sederholm, 1998), in line with the known acoustic dimorphism in these parameters (Titze, 1994), experimental (e.g. psychoacoustic) approaches, such as those used in the present thesis, are needed to directly explore the contribution of ΔF, as well as of other potential acoustic cues, to identify the sex of pre-pubertal child speakers.

From Sex Dimorphism to Gender Expression

So far we have seen that F0 and ΔF provide listeners with reliable cues to a speaker's sex, and that these acoustic differences largely originate from differences in the size of the vocal apparatus, and more generally, overall body size between males and females. The relationship between acoustic output, vocal apparatus and overall body size is common to many vertebrates: larger species (and within species, age and sex classes with larger individuals) are typically characterised by lower frequencies than smaller ones (dogs: Riede & Fitch, 1999; primates: Fitch, 1997; Lieberman, Klatt & Wilson, 1969; Owren, 1990; red deer: Reby & McComb, 2003). Stemming from these observations, the "size" or "frequency" code hypothesis posits that, within their anatomical constraints, callers may have evolved the ability to dynamically modify the source- and filter-related frequency components known to encode honest information on size and associated secondary meanings (e.g. physical dominance) through their vocalisations (Ohala, 1984). Indeed, in species where F0 and ΔF correlate negatively with body size, callers dynamically lower these acoustic traits in fighting contests in order to exaggerate the acoustic expression of their body size (sounding "bigger"), thus projecting greater physical strength or aggression (F0: male white-lipped frogs: Lopez et al*.,* 1988; ΔF: red-deer stags: Reby et al., 2005). Similarly, human males voices with lower frequency components (fundamental frequency and resonance frequencies) are typically seen as more attractive and dominant (Collins, 2000; Feinberg, DeBruine, Jones & Little, 2008; Feinberg, Jones, Little, Burt & Perrett, 2005; Hodges-Simeon, Gaulin & Puts, 2011Oguchi & Kikuchi, 1997; Puts, 2005; Puts et al., 2006), and human males have been found to lower their pitch when speaking to rivals whom they perceive as less physically dominant than them (Puts et al., 2006).

The "Gender Code" hypothesis

Recent studies in human speech have shown that adjustments of the same sexually selected voice cues (F0 and ΔF) are not restricted to the expression of sex, body size or physical dominance. More specifically, psychoacoustic studies suggest that F0 and ΔF are cues to "gender", encompassing the socially constructed roles and relationships, personality traits, attitudes, behaviours, values, relative power and influence that society ascribes to the two sexes on a differential basis (Money, 1955). For example, increasing F0 and formant frequencies can convey "feminine" traits (traits that are typically attributed to females), such as "femininity", "politeness", "modesty" and "vulnerability", and conversely, lowering of the frequency components of the voice can convey masculine traits (traits that are typically attributed to males) such as "masculinity", "competence", "social dominance" and "assertiveness" (Chuenwattanapranithi, Thipakorn & Maneewongvatana, 2008; Feinberg et al., 2005; Feinberg et al., 2008; Pisanski, Mishra & Rendall, 2012; Pisanski & Rendall, 2011; Puts et al., 2006; Simmons, Peters & Rhodes, 2011). Further, acoustic variation in F0 and ΔF exceeds biological differences between and within the sexes, as exemplified by the previously mentioned differences between pre-pubertal boys' and girls' voices, suggesting that some of this variation is under behavioural control.

On the basis of these observations, I propose to transpose Ohala's "size code" to the interpretation of gender-related variation in the human voice, making the prediction that a substantial proportion of vocal diversity can be explained by vocal gestures originating from a conventionalised use of primary, biologically-determined acoustic cues to sex. In other words, whereas Ohala's "size code" (1984) uses size-related variation of the vocal apparatus (and overall body) to explain the vocal expression of size-related secondary meanings, such as physical dominance and aggressiveness, the "gender code" uses sex-related variation to explain voice cues to gender and related attributes. More specifically, consistent with the view that human sexuality is socially constructed ("gender") as well as biologically determined, I argue that individuals perform adjustments of their primarily sexually dimorphic voice cues F0 and ΔF along a gender continuum, in order to alter their perceived gender and related attributes (masculinity and femininity), and in line with a variety of internal (e.g. speaker emotional state) and external (e.g. social) contexts.

Preliminary Evidence for the Existence of the "Gender Code"

In line with the "size code" hypothesis, three predictions must be satisfied for the "gender code" to exist: (i) F0 and ΔF are "honest" cues to sex; (ii) speakers can control F0 and ΔF within the given anatomical constraints; (iii) speakers perform global adjustments of F0 and ΔF in order to downplay or accentuate gender attributes depending on the desired outcome of the interaction and these adjustments are relevant to listeners (adapted from Ohala, 1996). So far, I have evidenced the first prediction by relating the acoustic dimorphism in F0 and ΔF to underlying biological dimorphisms, as well as the salience of F0 and ΔF in judging whether a speaker is male or female from their voice only. Different fields of voice-related research can gather preliminary evidence for the other two predictions. It is to this evidence that I now turn.

Variation of voice gender expression and acting. Humans' capacity for complex vocal imitation has long been recognised as playing a crucial role in the evolution of language. For example, several studies have shown that human infants spontaneously acquire novel sound patterns by hearing the vocalisations of adults and mimicking them (Menn & Ratner, 1999). Besides its relevance to language learning, vocal imitation can also be used to mimic another speaker's voice both at an acoustic and perceptual level (Zetterholm, 2006), as exemplified by studies on voice imitations of professional impersonators. In relation to gender imitation, female impersonations by male actors in Chinese theatre (Tian, 2000) and drag acts (exaggerated personifications of opposite gender roles – Koistra, 1999) indicate that actors can change their voices to sound more "female-like". In addition to showing that speakers can change their voices to vary the expression of their gender, this literature points at voice imitation as a methodological tool to find out which features a voice impersonator picks out in the target voice and which features in the human voice are not changed. This methodological tool underpins the production studies in Chapter 4, where I asked children and adults to masculinise or feminise their voices to assess whether they spontaneously modify F0 and ΔF in order to vary their gender expression. It is also used in Chapter 5, where I looked at the voice features that actors might shift in order to project a stereotypical image of gay characters as “feminine”.

Variation of voice gender expression in real life: expression of gender identity and sexual orientation. I will argue, however, that speakers not only use the “gender code” in acting contexts, but also (consciously or unconsciously) in real life. One obvious example relates to individuals who were raised as males, but self-identify as females. For male-to-female transsexuals, hormone treatment does not alter the adult male vocal mechanism and thus voice change must be achieved behaviourally (Holmberg, Hillman, Hammarberg, Södersten & Doyle, 2001). Indeed, voice therapy can effectively raise speakers’ F0 (e.g. by increasing vibration and tension of the vocal folds) and ΔF (e.g. by lip spreading) towards female values in order to achieve a female-sounding voice (Titze, 1994; Wolfe, Ratusnik, Smith & Northrop, 1990; Carew, Dacakis, & Oates, 2007). However, preliminary evidence indicates that the relevance of the “gender code” may not be limited to transsexual voices. Indeed, research on sexual orientation and the voice suggests that voice control underlies the “gay speech” style displayed by some homosexual speakers and which is characterised by partial shifts in voice frequencies resulting in broadly feminised speech in gay men and broadly masculinised speech in lesbian women (Chi-kuk, 2007; Pierrhumbert, Bent, Munson, Bradlow & Bailey, 2004; Baeck, Corthals & Borsel, 2011; Rendall, Vasey & McKenzie, 2008). The “gender code” seems also perceptually relevant to listeners when assessing sexual orientation of speakers (e.g. listeners typically rate more masculine voices as more “straight” sounding than feminine voices in men, and more “gay” sounding in women: Munson, 2007). The link between the “gender code” and voice stereotypes e.g. sexual orientation, masculinity and femininity is further explored in Chapter 5.

Variation of voice gender expression in real life: expression of gender roles. Research on transsexual and homosexual voices suggests that while the "gender code" is used in everyday life, it may only concern a minority of individuals. I will argue, however, that its use is more ubiquitous. Adjustments of frequency components appear to be strongly interlinked to gender and gender roles. Masculine (lower) fundamental and formant frequencies increase perceptions of masculinity (DeBruine et al., 2010; Feinberg et al., 2005; Feinberg et al., 2008; Pisanski et al., 2012; Pisanski & Rendall, 2011; Simmons et al., 2011), physical strength (Sell et al., 2010; Puts, Apicella & Cárdenas, 2011), dominance (Borkowska & Pawlowski, 2010; Vukovic et al., 2011; Wolff & Puts, 2010;), competence (Klofstad, Anderson & Peters, 2012), leadership (Anderson & Klofstad, 2012; Tigue, Borak, O'Connor, Schandl & Feinberg, 2012), and social status (Stanford, Gregory, & Gallagher, 2002) among men. By contrast, naturally or artificially higher-pitched, higher-resonance voices in women are perceived as more feminine (Feinberg et al., 2008), polite (Loveday, 1981; Ohara, 1999), submissive (Hall, Irish, Roter, Ehrlich, & Miller, 1994), less competent (Montepare & Zebrowitz-McArthur, 1987), warmer (Berry, 1992) and modest (Van Bezooijen, 1995) than lower-pitched, lower-resonance voices. Correspondingly, psychoacoustic studies have shown that men prefer vocal femininity in female voices (Feinberg et al., 2011; Fraccaro et al., 2010) while women prefer vocal masculinity in male voices (Pipitone & Galup, 2008). The salience of F0 and ΔF variation in listeners' gendered attributions of speakers' qualities suggests that individuals may dynamically modify F0 and ΔF to exaggerate the expression of these attributes conveyed by their vocalisations in line with specific social roles and contexts. From a production perspective, we have already seen that pre-pubertal boys speak with lower ΔF than girls (but same F0), despite the lack of overall differences in the vocal apparatus of the two sexes, leading several authors to suggest that the vocal expression of gender in children, like other types of children's gendered behaviour, may be linked to children internalising appropriate articulatory strategies, so that they not only "look" (e.g. in the way they interact with others, the activities they engage with), but also "sound" like a male or a female (Sachs et al., 1973). In adults, cross-language studies have also shown that between sexes, differences in F0 (Traunmüller & Eriksson, 1994) and formants (Johnson, 2006) exceed values predicted by body size alone, suggesting that variation in these cues may be partly learnt to project gender roles specific to one's culture (Johnson, 2006). Most strikingly, acoustic differences between Japanese men and women are reported to be greater than in other languages (Hiramoto, 2010), and are mainly attributed to women speaking with unusually high F0 (Loveday, 1981). In turn, these differences seem to reflect expectations in Japanese society, wherein women's use of high vocal pitch is strongly linked to the expression of socio-cultural expectations of femininity (Hiramoto, 2010), and its associated qualities such as dependency, modesty and weakness (Van Bezooijen, 1995).

Taken together, these studies suggest that children and adult speakers may use subtle shifts in F0 and ΔF to vary the expression of their gender and related attributes when complying with their own gender identity and with varying gendered roles within and across different communities. Investigating the use of gender code in these two populations (pre-pubertal children and adults) is the main objective of this thesis.

Research Questions and Thesis Outline: Summary

The main aim of the research is to explore speakers' behavioural control of sexually dimorphic voice cues in order to vary the expression of gender and related attributes, and the relationship between this acoustic variation and listeners' gendered perceptions of individuals (e.g. their gender, masculinity and femininity) within the "gender code" hypothesis. From a methodological perspective, the acoustic analyses and psychoacoustic manipulations of speech utterances will be carried out within the "source-filter" theory of speech production (Fant, 1960). From a theoretical perspective, the "gender code" applies the principles of the "size code" hypothesis to understanding the covariation of acoustic characteristics with gender variation. The "gender code" is schematically represented in Figure 1.2.

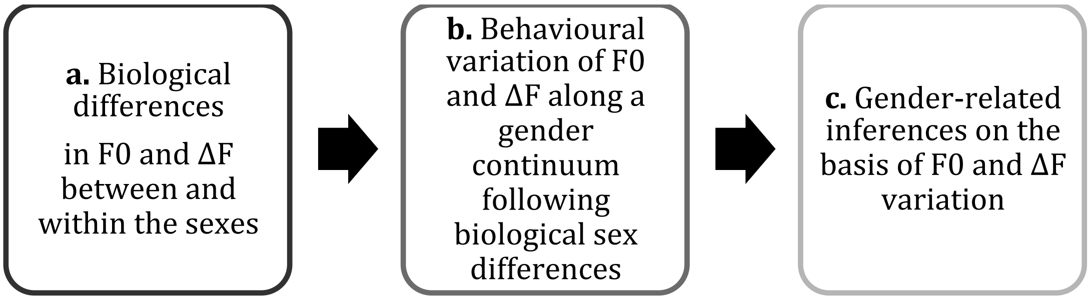

Figure 1.2. The "gender code" hypothesis. Inter and intra-individual gender differences in F0 and ΔF primarily reflect biological differences affecting the shape and dimensions of the vocal apparatus (a). However, speakers can dynamically modify these cues along this sex dimorphism to exaggerate or downplay the gender attributes of their voice (b). Listeners attend to F0 and ΔF variation and use it to make gendered attributions of speakers (c).

More specifically, using a combination of production and perception experiments with pre-pubertal children and adults, I aim to explore the following questions:

Question 1. How does the natural variation in the gender-related acoustic cues relate to anatomical and biological differences? (Chapter 3 – Studies 1 and 3)
Question 2. What is the perceptual relevance of this variation in terms of listeners' gendered attributions of speakers? (Chapter 3 – Studies 2 and 3)
Question 3. Can individuals control fundamental and formant frequencies in order to vary the expression of gender, masculinity and femininity of their voice, and does the acoustic co-variation of these parameters occur along the existent sex dimorphism? (Chapter 4 – Studies 4 and 5)
Question 4. Are speakers aware of what voice and articulatory gestures they use to vary the gender expression in their voice? (Chapter 4 – Studies 4 and 5)
Question 5. What is the perceptual relevance of these gestures? (Chapter 4 – Study 6)
Question 6. How does the "gender code" interact with cultural stereotypes in the expression and perception of gender attributes and sexual orientation? (Chapter 5 – Studies 7 and 8)

Chapter Two provides details of the materials and methods used to collect and analyse the data for my thesis.

Chapter Three explores Questions 1 and 2, by looking at the extent to which biological factors contribute to the diversity in voice gender expression, and to which naturalistic variation in F0 and ΔF is attended to by listeners and influences their gendered attributions of speakers. *Study 1* compares published anatomical measurements of vocal tract lengths with acoustic data on vocal tract resonances from childhood to adulthood. I predict the confirmation of a mismatch between sex differences in vocal tract length and sex differences in formant spacing (ΔF), particularly prior to puberty. *Study 2* uses a psychoacoustic approach to investigate the perceptual relevance of the naturalistic variation in pre-pubertal ΔF to gendered attributions made by adult listeners. I predict that ΔF will affect listeners' characterisations of speakers' gender (e.g. how masculine a child sounds) as well as their sex (e.g. whether the child is male or female). *Study 3* turns the focus on adult voices by exploring the extent to which sexually dimorphic cues in adulthood (F0 and ΔF) mediate between men's physical masculinity (e.g. body height and testosterone) and their perceived masculinity (as rated by women listeners) via path analysis. As for children's voices in the previous two studies, I predict that the observed acoustic and perceptual variation in voice gender of adults (at least in men) will only be partly explained by observed biological factors (e.g. taller, higher-testosterone individuals will have lower-pitched, more resonant voices).

Chapter Four focuses on Questions 3, 4 and 5. The first two studies use an imitation paradigm to investigate the spontaneous ability in pre-pubertal children (*Study 4*) and adults (*Study 5*) to masculinise or feminise their voices, as well as the acoustic (F0 and ΔF adjustments) and behavioural (e.g. lip movements to vary vocal tract length) correlates underlying such variation. I predict that, within their anatomical constraints, speakers will vary the sexually dimorphic cues of their voices (ΔF in pre-pubertal and adult speakers and F0 in adult speakers) to express gender and related attributes by lowering those parameters to masculinise their voices and by raising the same cues to feminise them. Speakers' awareness of the vocal and articulatory gestures involved is also explored via questionnaires. I predict that adults, and to a lesser extent children, will be aware of the perceptual output of their vocal gestures (e.g. sounding "lower" and "deeper" to sound more masculine). I also predict some awareness of related articulatory adjustments underlying such vocal gestures, and in particular of vocal tract length adjustments (via lip and laryngeal movements) aimed at varying formant dispersion, therefore feminising or masculinising one's voice. *Study 6* investigates the perceptual relevance of voice adjustments by asking listeners to rate the masculinised and feminised voices from the previous two studies. I predict that listeners will be attentive to voice frequency shifts when making gendered characterisations of speakers from their voices: speakers will receive highest masculinity ratings when they masculinise their voices (by lowering F0 and ΔF) and lowest masculinity ratings when feminising them (by raising F0 and ΔF).

Chapter Five explores Question 6, the variation in voice gender expression in relation to different social roles and contexts, by focusing on the interaction of the "gender code" and cultural stereotypes. *Study 7* looks at listeners' spontaneous expectations on pre-pubertal children's voices in relation to gender-stereotypical information (e.g. would a child that plays with dolls also sound feminine?). I predict that pre-pubertal and adult listeners will spontaneously associate resonance variation in children, as previously shown in adults, with gender-stereotypical attributions. *Study 8* looks at whether speakers dynamically modify their voice gender in response to stereotypes by testing the hypothesis that actors playing homosexual roles feminise their voices by raising their F0 and ΔF, thus reproducing in the auditory dimension stereotypical notions which attribute feminine characteristics to male homosexuality.

Chapter Six provides a summary of the results and specific directions for future research in relation to the six research questions underlying this thesis.

Chapter Seven discusses the importance of the "gender code" in providing a unifying framework for understanding variation in voice gender expression by linking evolutionary and social perspectives. The chapter ends by highlighting the potential impact of this research beyond the scientific community directly involved in the study of human vocal communication.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice, 13*(3), 424–446.

Anderson, R. C., & Klofstad, C. A. (2012). Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles. *PloS One, 7*(12), e51216.

Andersson, M. (1994). *Sexual selection*. NJ:Princeton University.

Aronovitch, C. D. (1976). The Voice of Personality: Stereotyped Judgments and their Relation to Voice Quality and Sex of Speaker. *The Journal of Social Psychology, 99*(2), 207–220.

Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer and H. Giles (Eds.) *Social markers in speech* (pp.147-209). Cambridge: Cambridge University Press.

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America, 106*(2), 1054–1063.

Baeck, H., Corthals, P., & Borsel, J. V. (2011). Pitch characteristics of homosexual males. *Journal of Voice, 25*(5), e211–e214.

Bahrick, L. E., Netto, D., & Hernandez-Keif, M. (1998). Intermodal perception of adult and child faces and voices by infants. Child Development, 69(5), 1263–1275.

Baker, S., Weinrich, B., Bevington, M., Schroth, K., & Schroeder, E. (2008). The effect of task type on fundamental frequency in children. *International Journal of Pediatric Otorhinolaryngology, 72*(6), 885–889.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America, 69*(1), 231–238.

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics.*

Bennett, S., & Weinberg, B. (1979). Sexual characteristics of preadolescent childrens' voices. *The Journal of the Acoustical Society of America, 65*(1), 179–189.

Berry, D. S. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior, 16*(1), 41–54.

Britto, A. I., & Doyle, P. C. (1990). A comparison of habitual and derived optimal voice fundamental frequency values in normal young adult speakers. *The Journal of Speech and Hearing Disorders, 55*(3), 476–484.

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences, 273*(1582), 83–89.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour, 82*(1), 55–59.

Britto, A. I., & Doyle, P. C. (1990). A comparison of habitual and derived optimal voice fundamental frequency values in normal young adult speakers. *The Journal of Speech and Hearing Disorders, 55*(3), 476–484.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America, 97*(4), 2603–2606.

Calhoun, K. H., & Eibling, D. E. (2013). Geriatric Otolaryngology. CRC Press.

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice, 21*(5), 591–603.

Charlton, B. D., Keating, J. L., Kersey, D., Rengui, L., Huang, Y., & Swaisgood, R. R. (2011). Vocal cues to male androgen levels in giant pandas. *Biology Letters, 7*(1), 71–74.

Charlton, B. D., Reby, D., & McComb, K. (2007). Female red deer prefer the roars of larger males. *Biology Letters, 3*(4), 382–385.

Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2009). The information content of giant panda,< i> Ailuropoda melanoleuca</i>, bleats: acoustic cues to sex, age and size. *Animal Behaviour, 78*(4), 893–898.

Charrier, I., Ahonen, H., & Harcourt, R. G. (2011). What makes an Australian sea lion (Neophoca cinerea) male's bark threatening? *Journal of Comparative Psychology, 125*(4), 385–392.

Chi-kuk, H. A. U. (2007). A Study on Acoustic Characteristics of Cantonese Gay Speech and Perceived Sexual Orientations. *The Chinese University of Hong Kong.*

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code. A perceptual investigation. *Phonetica, 65*(4), 210–230.

Coleman, R. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research, 19*(1).

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*(6), 773–780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*(5), 997–1004.

DeBruine, L., Jones, B. C., Frederick, D. A., Haselton, M. G., Penton-Voak, I. S., & Perrett, D. I. (2010). Evidence for Menstrual Cycle Shifts in Women's Preferences for Masculinity: A Response to Harris (in press) Menstrual Cycle and Facial Preferences Reconsidered. *Evolutionary Psychology, 8*(4).

Deliyski, S. A. X. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology, 27*(2), 159–168.

Dixson, A. F. (2012). Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans. Oxford University Press.

Eguchi, S., & Hirsh, I. J. (1968). Development of speech sounds in children. *Acta Oto-Laryngologica. Supplementum, 257*, 1–51.

Fant G (1960) Acoustic Theory of Speech Production. The Hague: Mouton & Co.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior, 29*(4), 233–241.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., O'Connor, J. J. M., Tigue, C. C., & Borak, D. J. (2011). Integrating fundamental and formant frequencies in women's preferences for men's voices. *Behavioral Ecology, 22*(6), 1320–1325.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561–568.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*(2), 1213–1222.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511.

Fitch, W. T. & Hauser, M. D. (2003). In A. M., Simmons, A. N., Popper, R. R. Fay, &

H. C. Gerhardt (Eds). Acoustic communication (Vol. 16, p.103). Springer.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. Proceedings of the Royal Society of London. *Series B: Biological Sciences, 268*(1477), 1669–1675.

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology, 8*(3).

Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldaschev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (Gazella subgutturosa). *Journal of Anatomy, 218*(5), 566–585.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology, 6*(4), 219–225.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice, 19*(4), 544–554.

González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics, 32*(2), 277–287.

Graddol, D., & Swann, J. (1989). *Gender voices*. Cambridge: Basil Blackwell Cambridge MA.

Greisbach, R. (2007). Estimation of speaker height from formant frequencies. *International Journal of Speech Language and the Law, 6*(2), 265–277.

Hall, J. A., Irish, J. T., Roter, D. L., Ehrlich, C. M., & Miller, L. H. (1994). Gender in medical encounters: An analysis of physician and patient communication in a primary care setting. *Health Psychology, 13*(5), 384–392.

Hasek, C. S., Singh, S., & Murry, T. (1980). Acoustic attributes of preadolescent voices. *The Journal of the Acoustical Society of America, 68*(5), 1262–1265.

Henton, C. (1995). Pitch dynamism in female and male speech. *Language & Communication, 15*(1), 43–61.

Hewlett, N., & Beck, J. M. (2013). *An Introduction to the Science of Phonetics*. New York: Routledge.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, &*

Psychophysics, 71(5), 1150–1166.

Hiramoto, M. (2010). Utterance final position and projection of femininity in Japanese. *Gender & Language, 4*(1).

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2011). Voice correlates of mating success in men: examining ‘contests’ versus ‘mate choice’ modes of sexual selection. *Archives of Sexual Behavior, 40*(3), 551–557.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America, 96*(5), 2646–2654.

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research, 15*(1), 155.

Holmberg, E. B., Hillman, R. E., Hammarberg, B., Södersten, M., & Doyle, P. (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice, 15*(3), 395–412.

Honjo, I., & Isshiki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Archives of Otolaryngology—Head & Neck Surgery, 106*(3), 149.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America, 106*(3), 1532–1542.

Hummert, M. L., Mazloff, D., & Henry, C. (1999). Vocal characteristics of older adults and stereotyping. *Journal of Nonverbal Behavior, 23*(2), 111–132.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*(4), 485–499.

Kahane, J. C. (1982). Growth of the human prepubertal and pubertal larynx. *Journal of Speech, Language, and Hearing Research, 25*(3), 446–455.

Kent, R. D. (2002). Acoustic analysis of speech. Singular Publishing Group.

Kent, R. D., & Vorperian, H. K. (1995). *Development of the craniofacial-oral-laryngeal anatomy.* San Diego: Singular Publishing Group.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences, 279*(1738), 2698–2704.

Kooistra, L. J. (1999). Cross-Dressing Confessions: Men Confessing as Women. Confessional Politics: *Women's Sexual Self-Representations in Life Writing and Popular Media, 167*.

Krook, M. I. (1988). Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatrica et Logopaedica, 40*(2), 82–90.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America, 100*(4), 2425–2438.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America, 59*(3), 675–678.

Lee, S., & Iverson, G. K. (2009). Vowel development in English and Korean: Similarities and differences in linguistic and non-linguistic factors. *Speech Communication, 51*(8), 684–694.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105,* 1455.

Lieberman, P. (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge University Press.

Lieberman, P., Klatt, D. H., & Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science, 164*(3884), 1185–1187.

Linders, B., Massa, G. G., Boersma, B., & Dejonckere, P. H. (1995). Fundamental voice frequency and jitter in girls and boys measured with electroglottography: influence of age and height. *International Journal of Pediatric Otorhinolaryngology, 33*(1), 61–65.

Linville, S. E. (2004). The aging voice. *The ASHA Leader, 9*(19), 12–21.

Lopez, P. T., Narins, P. M., Lewis, E. R., & Moore, S. W. (1988). Acoustically induced call modification in the white-lipped frog,< i> Leptodactylus albilabris</i>. *Animal Behaviour, 36*(5), 1295–1308.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech, 24*(1), 71–89.

Maltry, M., & Tucker, K. (2002). Female Fem(me)ininities. *Journal of Lesbian Studies, 6(*2), 89–102.

McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (Dama dama) to lower their formant frequencies. *Journal of Zoology, 270*(2), 340–345.

Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice, 10*(1), 59–66.

Money, J. (1955). Hermaphroditism, gender and precocity in hyperadrenocorticism: psychologic findings. Bulletin of the Johns Hopkins Hospital, 96(6), 253–264.

Menn, L., & Ratner, N. B. (1999). *Methods for Studying Language Production* (pp. 27-41). Psychology Press.

Montepare, J. M., & Zebrowitz-McArthur, L. (1987). Perceptions of adults with childlike voices in two cultures. *Journal of Experimental Social Psychology, 23*(4), 331–349.

Mulac, A., & Giles, H. (1996). 'Your're Only As Old As You Sound': Perceived Vocal Age and Social Meanings. *Health Communication, 8*(3), 199–215.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech, 50*(1), 125–142.

Nanda, S. (1999). Gender Diversity: Crosscultural Variations. Waveland Press.

Oguchi, T., & Kikuchi, H. (1997). Voice and interpersonal attraction. *Japanese Psychological Research, 39*(1), 56–61.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*(1), 1–16.

Ohala, J. J. (1996). Ethological theory and the expression of emotion in the voice. In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International

Conference on (Vol. 3, pp. 1812–1815). IEEE.

Ohara, Y. (1999). Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English. In *Wahrnehmung und Herstellung von Geschlecht* (pp. 105–116). Springer.

Owren, M. J. (1990). Acoustic classification of alarm calls by vervet monkeys (< em> Cercopithecus aethiops) and humans (< em> Homo sapiens): II. Synthetic calls. *Journal of Comparative Psychology, 104*(1), 29.

Owren, M. J., Berkowitz, M., & Bachorowski, J.-A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics, 69*(6), 930–941.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America, 109,* 2988.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America, 116,* 1905.

Pipitone, R. N., & Gallup Jr, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior, 29*(4), 268–274.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior.*

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*(4), 2201–2212.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior, 26*(5), 388–397.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601–609.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior,* *27*(4), 283–296.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*(3), 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. Tshi., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences, 272*(1566), 941–947

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America, 117*(2), 944–955.

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The Queen's English: An alternative, biosocial hypothesis for the distinctive features of 'gay speech'. *Archives of Sexual Behavior, 37*(1), 188–204.

Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris). *Journal of Experimental Biology, 202*(20), 2859–2867.

Russell, A., Penny, L., & Pemberton, C. (1995). Speaking fundamental frequency changes over time in women: a longitudinal study. *Journal of Speech and Hearing Research, 38*(1), 101.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Sederholm, E. (1998). Perception of gender in ten-year-old children's voices. *Logopedics Phonatrics Vocology, 23*(2), 65–68.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Rueden, C. von, … Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences,* rspb20100769.

Simmons, L. W., Peters, M., & Rhodes, G. (2011). Low Pitched Voices Are Perceived as Masculine and Attractive but Do They Predict Semen Quality in Men*? PLoS*

ONE, 6(12), e29271.

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research, 296,* 131–140.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and agea). *The Journal of the Acoustical Society of America, 118*(5), 3177–3186.

Stanford W., Gregory JR., & Gallagher, T. J. (2002). Spectral Analysis of Candidates' Nonverbal Vocal Communication: Predicting U.S. Presidential Election Outcomes. *Social Psychology Quarterly, 65*(3), 298.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221–236.

Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog,< i> Canis familiaris</i>, growls. *Animal Behaviour, 79*(1), 205–210.

Tian, M. (2000). Male Dan: The Paradox of Sex, Acting, and Perception of Female Impersonation in Traditional Chinese Theatre. *Asian Theatre Journal, 17*(1), 78–97.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33*(3), 210–216.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324–333.

Traunmüller, H., & Eriksson, A. (1994). The frequency range of the voice fundamental in the speech of male and female adults. Manuscript, Department of Linguistics, University of Stockholm,(Accessed May 8 2004)

Trollinger, V. L. (2003). Relationships between pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender in American English-speaking preschool children. *Journal of Research in Music Education, 51*(1), 78–94.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech, 38*(3), 253–265.

Van Rie, J., & Van Bezooijen, R. (1995). Perceptual characteristics of voice quality in dutch males and females from 9 to 85 years. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 290–293).

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study a). *The Journal of the Acoustical Society of America, 125*(3), 1666–1678.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research, 54*(4), 995–1010.

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L. M., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology, 102*(1), 37–48.

Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development, 14*(2), 131–142.

Weinberg, B., & Bennett, S. (2005). Speaker Sex Recognition of 5- and 6-Year-Old Children's Voices. *The Journal of the Acoustical Society of America, 50*(4B), 1210–1213.

Welch, G. F., & Howard, D. M. (2002). Gendered voice in the cathedral choir. *Psychology of Music, 30*(1), 102–120.

Whiteside, S. P. (1998). The identification of a speaker's sex from synthesized vowels. *Perceptual and Motor Skills, 87*(2), 595–600.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America, 110,* 464 (p.178).

Whiteside, S. P., & Hodgson, C. (1999). Acoustic characteristics in 6-10-year-old children's voices: some preliminary findings. *Logopedics Phonatrics Vocology, 24*(1), 6–13.

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders, 55*(1), 43–50.

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology, 64*(10), 1673–1683.

Wyman, M. T., Mooring, M. S., McCowan, B., Penedo, M. C. T., Reby, D., & Hart, L. A. (2012). Acoustic cues to size and quality in the vocalizations of male North American bison, Bison bison. *Animal Behaviour, 84*(6), 1381–1391.

Yamada, N., Hakoda, Y., Yuda, E., & Kusuhara, A. (2000). Verification of impression of voice in relation to occupational categories. *Psychological Reports, 86*(3c), 1249–1263.

Yogo, Y., Ando, M., Hashi, A., Tsutsui, S., & Yamada, N. (2000). Judgments of emotion by nurses and students given double-bind information on a patient's tone of voice and message content. *Perceptual and Motor Skills, 90*(3 Pt 1), 855–863.

Zetterholm, E. (2006). Same speaker–different voices. A study of one impersonator and some of his different imitations. In *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 70–75).

Chapter 2: Materials and Methods

Speakers

Adult speakers from Studies 3, 5 and 6 were recruited from students and faculty members at the University of Sussex and individually audio-recorded in a sound booth on campus. Study 8 focused on actors' recordings, which were taken from selected TV shows and interviews. Child speakers from Studies 2, 4, 6 and 7 were recruited from Hurst Pre-prep School (Hurstpierpoint), Lewes YMCA (Lewes), Middle Street Primary School (Brighton), Holy Trinity CE Primary School (Horsham) and through emails and posters distributed to staff and students at the University of Sussex. Children were individually audio-recorded in a quiet room at their school or in a sound booth on campus. Child and adult speakers had no history of hearing or speech impediments and were all native speakers of British English. For all speakers, body height, body weight, age and sex were recorded prior to their experimental session.

Listeners

Adult listeners were recruited from students and faculty members at the University of Sussex (Studies 2, 3, 6, 7). Child listeners were recruited from Holy Trinity CE Primary School (Horsham) and through emails and posters distributed to staff and students at the University of Sussex (Study 7). All listeners had no history of hearing or speech impairments and were speakers of English.

Sex and gender ratings

Sex attributions of child speakers from their voices ("Please identify the sex of the speaker") were made by choosing between "male" or "female" options (Study 2). Gender ratings of child speakers from their voices ("Rate the voice of the speaker on a scale of 1 to 7") were made on a seven-point Likert scale (1= masculine boy, 2 = boy, 3 = feminine boy, 4 = neutral, 5 = masculine girl, 6 = girl, 7= feminine girl) (Studies 2, 6, 7).

Vocal masculinity of adult speakers ("how masculine does the speaker sound?") was assessed on a seven-point Likert scale from 1 (not at all masculine) to 7 (very masculine) (Studies 3 and 5). The definition of masculinity was left open in order to gauge listeners' spontaneous assessments of speakers along this gender dimension, without any prior cuing.

Body measurements

Body height was recorded to the nearest 0.1cm using a freestanding *Seca Leicester Stadiometer*, after participants took their shoes off and stood with their shoulders flush to the stick and their heads level and oriented forward (Studies 3, 4, 5). Body weight was measured to the nearest 0.1kg using *PS250* veterinary or *Hanson* floor scales (Studies 4 and 5).

Testosterone measurements

Testosterone was collected from male speakers in Study 3. Testosterone is a steroid hormone synthesised in the testes in males, the ovaries in females, and adrenal glands in both sexes (Salimetrics, 2012). Testosterone levels can be measured by taking saliva samples, given that the majority of testosterone in saliva is not protein-bound and is not affected by salivary flow rate or salivary enzymes (Salimetrics, 2012). Speakers' saliva was collected with the *Salimetrics* Salivary *Testosterone* Enzyme Immunoassay Kit. As testosterone exhibits a diurnal rhythm, with highest levels in the morning and a nadir around midnight (Evans et al., 2008), testosterone samples were collected between 9 and 11 am to maximise comparability across participants. Samples were analysed to measure speakers' testosterone levels using immunoessay analysis by Salimetrics Ltd. The immunoassay analysis principle works by adding the testosterone in the sample (unlabelled analyte) and the testosterone linked to horseradish peroxidase (labelled analyte) to a microplate coated with rabbit antibodies to testosterone. The labelled and unlabelled analytes compete for the antibody binding sites and after incubation unbound components are washed away. The remaining labelled, bound analyste is measured by observing the signal (change in colour) resulting from the binding between the anti-bodies and the labelled analyte. The more testosterone in the sample, the more labelled analyte gets competed off and hence the amount of labelled, bound analyte is inversely proportional to the amount of testosterone in the sample: the lower the coloured signal, the more testosterone there is in the sample (Salimetrics, 2012).

Lip measurements

Lip measurements of adult speakers were taken in Study 5. Speakers' horizontal mouth corners and the upper and lower centre lips were marked using a black makeup pencil. Video recordings were taken using a *Sony HDR-TG3E handycam*. Still frames of speakers as they uttered vowels were selected from the video recordings via *Apple iMovie version 8.0.6* and the line drawing function in Adobe Illustrator CS5 was used to calculate vertical (Lip Openness - LO) and horizontal (Lip Spreading - LS) distances from the markers. Lip Ratio (LR) was also calculated from the two distances as LS / LO.

Speech corpora

A variety of speech materials were selected for acoustic analysis, including vowels, sentences, text extracts and running speech. Vowels (Studies 1 to 7) were extracted from a consonant-vowel-consonant (CVC) context to be free of any phonological or lexical restrictions (due to different encoding methods in different journals, vowels are represented by two systems in this thesis. See Table 2.1 for vowel symbols correspondences). In adults, the context chosen was /hVd/, as /h/ would be more transparent to coarticulation (Cox, 2002). In children I had to compromise by using words that children knew and found easy to understand and pronounce. The steady-state portion of the vowel (e.g. where formants are stable) was then used for acoustic analyses. Short text extracts of neutral content (Studies 3 and 5) were also selected as they are thought to be closer to sentence-based real-life phonation and therefore less prone to biases produced by the shortness and artificiality of the vowel task (Moon, Chung., Park, S., & Kim, 2012). Similarly, (equal-sized) segments were extracted from running speech (Studies 3 and 8), which was used to measure a more naturalistic response than read speech. In Study 3, speakers were asked to spontaneously describe the same object (a kettle) for one minute, ending with the statement "the object I have in front of me is a kettle", which was then selected for analysis. This allowed to elicit spontaneous speech and yet obtain the same phonetic data (LListerri, 1992).In Study 8, running speech was selected from video clips with no background noise (e.g. music, other people speaking), no crowded settings (e.g. office, bar) and no strong emotional content.

Table 2.1

International Phonetic Association IPA symbols and SUN-style for vowel stimuli

IPA	æ	ɛ	ʌ	ɜ	i	ɪ	ɒ	ʊ	u
SUN	ae	eh	ah	er	ih	iy	aa	uh	uy

Audio recordings

For adults, all recordings were obtained in a soundproofed room at the University of Sussex using a high-fidelity (*AKG Perception 220*) microphone, which was typically held at 30 cm from the participant's mouth and connected to a *Marantz PMD670*. Audio recordings of children were made using a high-fidelity *Shure SM94* microphone connected to a *Tascam DR07mkII* handheld recorder from a sound-attenuated room at the children's school and at the University of Sussex. Recordings were then transferred to a MAC mini (OS X v.10.6.6) into mono WAVE format with a bit rate 128 kbps and 44.1kHz sampling rate for acoustic analysis. To ensure sound quality, coughs, laughter and other non-speech noises were all removed from the samples. As acoustic analyses at the boundaries do not function properly at the beginning and end of the signal (Wood, 2003), 0.5s of silence were added at the beginning and end of each sample. Subsequently, samples were scaled in intensity to a 65dB level using the "Scale intensity" command in PRAAT, which multiplies the amplitude of the sound in such a way that its average (e.g. root-mean-square) intensity becomes the new average intensity.

Acoustic analyses

Both source- (fundamental frequency) and filter- (formant) related acoustic features were measured using the PRAAT freeware (versions 5.03 – 5.20, Boersma & Weenink, 2006, 2009, 2011). To assist in the estimates, I developed a custom, batch-processing PRAAT script from previous PRAAT scripts used to study animal vocalisations at the Mammal Vocal Communication Laboratory, University of Sussex. The script assigns a random identifier to each sample in order to ensure blind analysis.

It also allows the experimenter to set the analysis parameters and to adjust them after visually comparing the estimated frequencies with those tracked onto the spectrogram of the speech sound, in order to eliminate erroneous estimates.

Estimates of Fundamental Frequency Parameters

Fundamental Frequency (F0) parameters were calculated by the script using PRAAT autocorrelation algorithm ‘‘to Pitch’’ described in Boersma (1993). The script allows the experimenter to set time step, pitch floor and pitch ceiling parameters prior to the analysis. *Time step* is the measurement interval (frame duration) in seconds, which was typically set to 0.01s (100 pitch values per second – Boersma, 2003). *Pitch floor* determines the length of the analysis window and also represents the lowest fundamental frequency targeted within each sample (candidates below this frequency will not be recruited). The *pitch ceiling* represents the highest fundamental frequency targeted within each sample (candidates above the prescribed setting will be ignored). Different pitch floors and ceilings were used to reflect pitch variation according to task sex and age of speakers (e.g. adult males have lower ranges than adult females), but the expected fundamental frequency range was typically 30Hz-500Hz as it encompasses the natural range of variation in the human voice (Titze,1994). All the other parameters of the “to Pitch” command in PRAAT were left as default. The script then applies a low pass filter to smooth out rapid F0 changes within the specified range in the F0 contour before allowing the experimenter to visually check that the tracked F0 frequency points are correct. The mean value for F0 and its standard deviation (F0SD) are then extracted from the entire signal using the “get mean” command and “get standard deviation”, respectively. The script also calculates the coefficient of variation for F0 (F0CV), which is given by F0SD/F0mean. The coefficient of variation was used as an estimate of F0 variation as it provides a measure of the magnitude of F0 variation relative to the mean, reflecting the logarithmic perception between F0 and perceived pitch (Gaudio, 1994; Lee et al., 1999). For example, a F0SD of 200Hz for a mean F0 of 400Hz will be perceived as greater than a F0SD of 200Hz for a mean of 600Hz, because this is based on the ratio of the two frequencies ($F0CV_1$=400/200=2 and $F0CV_2$=600/400=1.5) rather than the absolute difference (200Hz). Therefore F0CV is a better estimate of F0 variation than its absolute estimate given by F0SD (Lee et al., 1999). Perceptually, a voice with lower F0CV has a more monotone quality than a voice with higher F0CV.

Estimates of Formant Frequencies Parameters

Formant frequency values were estimated by the script using PRAAT's Linear Predictive Coding (LPC: "To Formants (Burg)" command). LPC analysis is a widely used technique in estimating formant centre frequencies in both human speech and animal sounds (Fitch, 1997). LPC provides as output the coefficients of an *n*th-order all-pole digital filter whose frequency response best approximates the spectrum of the input signal. Formants are identified from the peaks in the spectrum. The script allows to set the following parameters prior to analysis: *window length*, which determines the size of the sound section PRAAT will examine to find the frequencies in the signal at that given moment; *number of formants*, which is the number of formants to be reported; and *maximum formant*, which is the upper threshold of the formant value to be tracked (frequencies above this limit are ignored). *Window lengths* were set between 0.0025s and 0.005s to produce a broadband spectrogram (the shorter the window, the broader the frequency bandwidth of the filter which performs the spectrographic analysis, resulting in less detailed frequency resolution: vowel formants are displayed while individual harmonics are not (Boersma, 2003)); *Number of formants* and *Maximum formant* were set for adult male speakers to reflect one formant in each 1000Hz band (e.g. 4 formants in 4000Hz (Wood 2003)). *Maximum formant* was adjusted by adding an additional 10-20% (e.g. 4 formants in 4400–4800Hz) for adult female speakers, and a 20-32% for children (e.g. 5 formants in 6000–6600Hz) to account for age- and sex-related differences in vocal tract size (Huber, Stathopoulos, Curione, Ash & Johnson, 1999). All LPC measurements were visually verified by superimposing the LPC-derived frequency response, showing as a continuous red trail through each formant, over the wideband spectrogram of the sound obtained by Fast Fourier Transform (FFT). The script allowed to manually adjust the analysis parameters to maximise formant estimation (red trail superimposing on the formant).

Formant spacing. The extracted frequency values of the first *n* formants (typically F1-F4) were then used to calculate formant spacing, defined as the "average distance between each adjacent pair of formants" (Fitch, 1997, p.1216):

(1) $\Delta F = F_{i+1} - F_i$

Because the vocal tract can be approximated to a uniform tube closed at one end (the glottis) and opened at the other (the mouth), the centre frequencies of the successive formants (F_1, F_2, .. F_i) generated by such a resonator are related to the length of the vocal tract by the equation:

(2) $F_i = \dfrac{(2i-1)c}{4VTL}$

where *c* is the speed of sound in air (approximated as 350m/s in the human vocal tract) and *VTL* is the length of the vocal tract. As a consequence, the spacing between any two consecutive formants in the frequency spectrum is constant and given by:

(3) $\Delta F = F_{i+1} - F_i = \dfrac{c}{2VTL}$

By replacing $\dfrac{c}{2VTL}$ in equation (2) with ΔF in equation (3), individual formant frequencies can be related to formant spacing ΔF by the equation:

(4) $F_i = \dfrac{2i-1}{2}\Delta F$

For each utterance, we can therefore estimate ΔF, the overall spacing of the formants, by seeking the best fit for equation (4) to the centre frequency of the first four formants, a method originally developed by Reby and McComb (2003) and illustrated in Figure 2.1. Because the length of the vocal tract is inversely proportional to the spacing between formant frequencies, individuals speaking with longer vocal tracts should have narrower overall formant frequency spacing. Crucially, changing the glottal and lip boundary conditions (whether the tube is opened or closed at the ends) would shift the entire pattern up or down in frequency, and thus the absolute frequencies of the formants, but would not change their spacing. A measure of formant spacing therefore overcomes the need to make assumptions about such conditions, providing an accurate estimate of vocal tract length (Riede & Fitch, 1999).

While the quarter-wave resonator is an accurate model for the unconstricted schwa sound, the production of other vowels involve constrictions in the oral tract, affecting individual formant values and therefore requiring more complex models. It is worth noting, however, that subject differences in VT length will still be reflected in ΔF differences for these vowels, and that averaging formants in connected speech leads to central vowel values, that reflect physical speaking (as opposed to resting) VTL quite accurately (Titze, 1994).

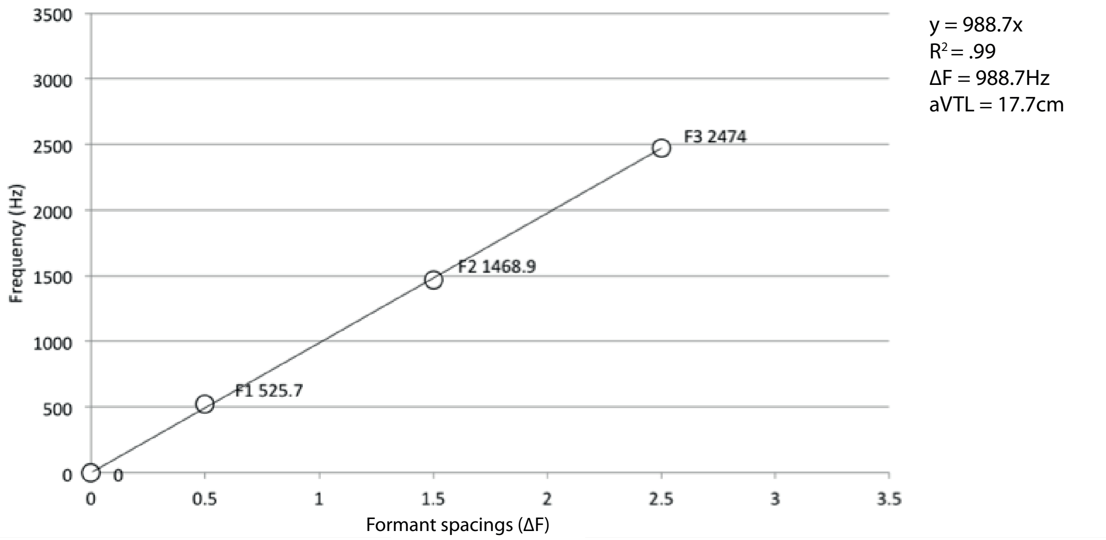

Figure 2.1. Illustration of the method used to estimate formants spacing. The observed frequency values of each formant (F1: 525.7Hz, F2: 1468.9Hz, F3: 2474Hz from Lee and colleagues (1999) on 28 male adults) are plotted against increments of the formant spacing as predicted by the vocal tract model. A linear regression line is subsequently fitted to the set of observed values, using an intercept equal to 0. ΔF is equal to the slope of the linear regression (ΔF = 988.7), which is fitted to the observed Fi values (method adapted from Reby and McComb (2003)).

Psychoacoustic experiments

The speech corpora for the psychoacoustic experiments (Studies 2, 3, 6, 7) consisted of natural and resynthesised voice stimuli from the production experiments. Natural voice stimuli were used to assess listeners' perceptions from individual variation as naturally occurring in the human voice (Studies 3, 6). Resynthesised stimuli were used to assess the individual contribution of acoustic parameters to listeners' perceptions (Studies 2, 7). In all psychoacoustic experiments, participants were sat in front of a laptop computer and wore *Dynamode dh-660mv* headsets. Sound volume was set to a comfortable level (default value of 65%), although participants were allowed to adjust the volume if needed by listening to a sound prior to the start of the experiment.

Stimuli were played back in a pseudo-random order using custom-based MCG scripts in Praat. Responses were saved by the scripts in a table on Praat and collected by exporting the table into Excel.

Acoustic Re-synthesis

Resynthesised stimuli were used in Studies 2 and 7 investigating the relative contribution of ΔF to the perception of gender. Acoustic analyses for each stimulus were run prior to resynthesis, to establish the factors needed to achieve the target ΔF values within the natural range of variation. All resynthesised stimuli were also subject to full acoustic analysis to check that resynthesis was successful. Resynthesis was implemented using PRAAT's Pitch Synchronous OverLap and Add (PSOLA) algorithm via the command "Convert: change gender". The algorithm enables the independent rescaling of individual acoustic parameters while leaving all other parameters unchanged, by dividing the signal into small overlapping signals (Zölzer et al, 2002), which are then resynthesised before overlap and adding. More specifically, linear scaling of formants is achieved by changing the duration of the signal in the resynthesis step so that some segments are either repeated multiple times (to increase the duration) or eliminated (to decrease the duration). As time scaling is the inverse of frequency scaling, if formant frequencies need to be increased by a factor α, every segment is shortened by a factor of 1/α (Zölzer et al, 2002). The segments are then recombined using the overlap add technique. An illustration of PSOLA resynthesis is provided in Figure 2.2.

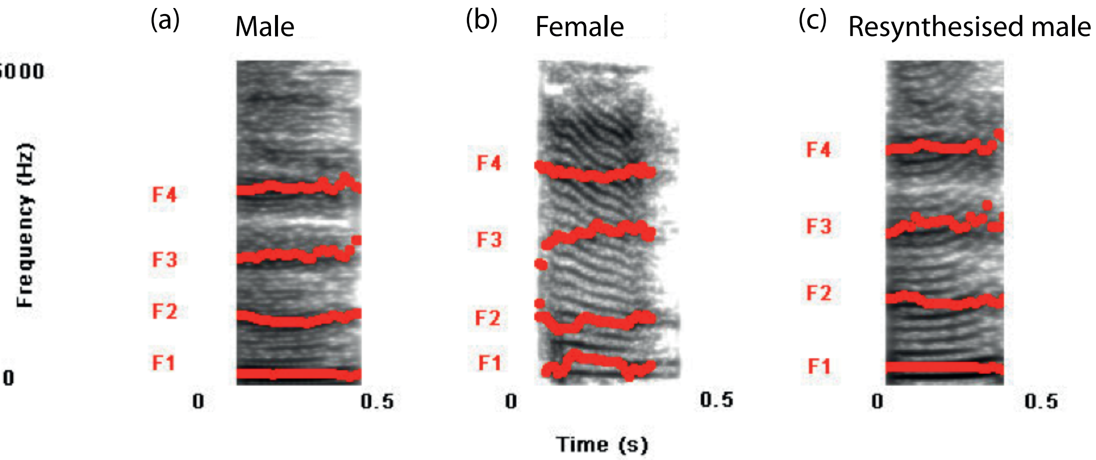

Figure 2.2. Spectrograms (a-c) to illustrate PSOLA resynthesis via the "Change gender" PRAAT command. Original male (a) and female (b) speech spectrum, linear modification of male spectrum (c) to approach female formant values (20% formant shift). The formants are labelled F1 – F4. (a) Male formant values - F1: 315Hz, F2: 1250Hz, F3: 2289Hz, F4: 3329Hz; (b) Female formant values - F1: 406Hz, F2: 1298Hz, F3: 2549Hz, F4: 3526Hz; (c) Resynthesised male to female F1: 363 Hz, F2: 1468Hz, F3: 2751Hz, F4: 4004Hz.

Ethical considerations

All studies received ethical approval by the Ethics Committee at the University of Sussex (authorization codes: DRVC0409, DRVC0709, DRVC0711).

All adult participants gave their informed consent in writing prior to taking part in the experiments. In addition, child participants gave their verbal informed consent. Guardian consent in writing was also obtained for all children.

References

Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer [Computer program v.5.0.3]. Retrieved from http://www.praat.org/

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer [Computer program v.5.1.0-20]. Retrieved from http://www.praat.org/

Boersma, P., & Weenink, D. (20011). Praat: doing phonetics by computer [Computer program v.5.2.17-20]. Retrieved from http://www.praat.org/

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-t-noise ratio of a samples sound. *IFA Proceedings,* 97-110.

Boersma, P. (2003). *Praat Manual*. Retrieved 20 October 2008, from: http://www.fon.hum.uva.nl/praat/manual/Manipulation.html

Cox, F. M. (2002). A description of the acoustic characteristics of/hVd/vowels in Australian English. *Australian Journal of Linguistics*.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4), 783–788.

Salimetrics Ltd. (2012). *Salivary Enzyme Immunoassay Kit*. Retrieved April 29, 2012 from: http://www.salimetrics.com/assets/documents/1-2402.pdf

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222.

Gaudio, R. P. (1994). Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men. *American Speech*, *69*(1), 30.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America*, *106*(3), 1532–1542.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Llisterri, J. (1992). Speaking styles in speech research. In *Workshop on Integrating Speech and Natural Language*. Retrieved from http://liceu.uab.cat/~joaquim/publicacions/SpeakingStyles_92.pdf

Moon, K. R., Chung, S. M., Park, H. S., & Kim, H. S. (2012). Materials of acoustic analysis: sustained vowel versus sentence. *Journal of voice: official journal of the Voice Foundation*, *26*(5), 563–565.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris). *Journal of Experimental Biology*, *202*(20), 2859–2867.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Wood, S. (2003). Praat for beginners [Manual]. Retrieved June 11, 2013, from http://person2.sol.lu.se/SidneyWood/praate/wavformedform.html

Zölzer, U., Amatriain, X., Arfib, D., Bonada, J., Poli, G. D., Dutilleux, P., … Todoroff, T. (2002). *DAFX - Digital Audio Effects*. John Wiley & Sons pp. 206-07.

Chapter 3:
Natural Variation in Sexually Dimorphic Cues Signals Gender

Summary

As a clear background to the “gender code” hypothesis, this chapter sets out to confirm that sexually dimorphic voice traits (ΔF from early childhood onwards, and F0 from puberty to adulthood) cue for speakers' gender characteristics, while highlighting that biological factors cannot fully explain the observed acoustic variation in voice gender. More specifically, the following questions will be explored:

Question 1. How does the natural variation in the gender-related acoustic cues relate to speakers’ anatomical and biological differences?
Question 2. What is the perceptual relevance of this variation in terms of listeners’ gendered attributions of speakers?

Study 1 investigates Question 1 by relating anatomic differences in vocal tract lengths of males and females throughout development (from age five to adulthood) to between-sex differences in its acoustic correlate, the overall spacing of vocal tract resonant frequencies (ΔF).

Summary of findings:

- The sex dimorphism in ΔF was largely attributable to sex differences in VTL. As the length of VT increased with both age and sex, ΔF decreased: children had shorter VTL, and thus higher ΔF, than their larger-bodied adult counterparts, while men, who have longer VTL than women (due to being bigger than women and subject to a male-specific descent of the larynx), also had lower ΔF than women.
- There was a mismatch between anatomical differences in VTL and acoustic differences in ΔF, particularly in pre-pubertal speakers, suggesting that some of the acoustic dimorphism in this parameter is linked to articulatory behaviours learnt through gender socialisation (e.g. boys learning to speak like a “man” and girls like a “woman”)

Study 2 investigates Question 2 in relation to pre-pubertal children's voices, by exploring whether variation in formant frequency spacing, which is sexually dimorphic since early childhood, affects their perceived gender.

Summary of findings:

- Small acoustic variation in ΔF (obtained by manipulating ΔF in increments of 2% within the natural range of children's voices) was salient and relevant to adult listeners when making sex (male, female) and gendered characterisations (masculinity, femininity) of child speakers from their voices
- The percentage of stimuli identified as female increased progressively as ΔF increased, following an S-shaped pattern for both boys' and girls' voices (no categorical perception of gender)
- Masculinity ratings decreased progressively as ΔF increases, following a linear curve for both boys' and girls' voices
- Boys' sex identification and gender rating curves were shifted bottom left compared to girls': boys voices were still perceived as more male and masculine than girls' despite similar ΔF values, suggesting that voice cues other than ΔF may also contribute to pre-pubertal gender signalling

Study 3 investigates both Question 1 and Question 2 in adult voices, by looking at how acoustic variation in sexually dimorphic voice cues (F0 and ΔF) mediates between biological and perceptual masculinity, highlighting the interdependence of these dimensions.

Summary of findings:

- Male speakers who were taller and had higher salivary testosterone levels, also had lower voice fundamental frequency (F0) and formant spacing (ΔF) and were in turn perceived as more masculine by women. However, variation in F0 and ΔF was only partly accounted for by speakers' hormonal (testosterone) and body size (height) profiles, suggesting that some of this acoustic variation may be behavioural in origin
- The relationship between testosterone and perceived masculinity was almost entirely mediated by F0, while the relationship between height and perceived masculinity was partially mediated by both F0 and ΔF

Study 1:

The Biological Dimorphism is not sufficient to Explain Gender Differences in the Voice: A Review and Re-analysis of Published MRI and Acoustic Data

Abstract

The human voice typically conveys cues to gender: men speak with lower voices (lower fundamental and formant frequencies) than women, and pre-pubertal boys speak with lower voices (lower formant frequencies) than girls. While these acoustic differences are mostly down to the size dimorphism (male > female) of the vocal apparatus, converging evidence suggests that the acoustic diversity of gender expression may also have a behavioural dimension. Here, we propose that, on top of the biologically determined sexual dimorphism, humans use a “gender code” consisting of vocal gestures to modulate the apparent gender of their voice e.g. lowering their voice frequencies to sound more masculine, and raising them to sound more feminine. To investigate this hypothesis, we review data on sex variation in resting anatomical Vocal Tract Length (MRI-VTL), and apparent vocal tract length or aVTL (an estimate of speaking anatomical VTL based on the overall spacing of formant frequencies). Our results confirm that between-sex differences in aVTL exceed differences in MRI-VTL throughout development, and in particular prior to puberty, where no sex differences in MRI-VTL are reported. The observed mismatch between acoustic and anatomical observations suggests that speakers may accentuate the perceived gender of their voice by dynamically varying their vocal tract length, thus affecting their formant frequencies: males tend to speak with lower ΔF, and females with higher ΔF than expected from their anatomical VTL. These gender-specific behaviours seem to play a particularly prominent role in the absence of anatomically determined sex differences (e.g. prior to puberty).

Introduction

The growing availability of Magnetic Resonance Imaging (MRI) has greatly facilitated investigations of the anatomical development of the human vocal tract. Using this technique, studies have shown that the vocal tract rapidly lengthens from 8 cm to about 10 cm between 0 and 24 months of age, and another centimetre in the following 12 months (Vorperian, Kent, Gentry & Yandell, 1999; Vorperian et al., 2005), due to the rapid descent of the hyoid bone and the larynx (Lieberman, McCarthy, Hiiemae & Palmer, 2001). MRI data have also revealed that the length of male and female vocal tracts follows a linear, gradual growth during the pre-pubertal period and post-pubertal periods, in line with the developmental trajectory in overall body size, while males' tracts undergo a disproportionate lengthening during male puberty (between ages of 11 and 15, Fitch & Giedd, 1999; Vorperian et al., 2009). The reasons behind this male-specific pubertal growth spurt in vocal tract length are well understood: under the influence of androgens, males grow about 7% taller than females (Gaulin & Boster, 1985), and are subject to a second laryngeal descent which differentially lengthens their pharynx by about 7mm on average (Fitch & Giedd, 1999). As a result, by the end of their sexual maturation (between 16 and 19 years of age) males develop vocal tracts that are 1.5-2 cm longer than females' on average (males' VTL: ~17cm, females' VTL: ~15 cm (Fitch & Giedd, 1999; Vorperian et al., 2009)).

The differential lengthening of males' and females' vocal tracts throughout development is accompanied by sex and age related differences in vocal tract resonances (formants), with longer tracts producing lower, more closely spaced formants (Fant, 1960). Tracking concomitant differences in vocal tract length, men's formant frequencies are lower and more narrowly spaced than women, while children speak with higher, more widely spaced formant frequencies than their adult counterparts (Lee, Potamianos, & Narayanan, 1999; Rendall, Vokey, & Nemeth, 2007). However, age- and sex- related growth patterns in vocal tract length do not always match corresponding changes in formant values. For example, as the vocal tract rapidly lengthens from infancy to two years of age, F3 decreases, while F1 and F2 remain stable (Gilbert, Robb & Chen, 1997; Robb, Chen & Gilbert, 1997; Vorperian et al., 1999). Moreover, while men have 20% longer tracts than women, female formant frequencies cannot be entirely reduced to male values by a simple scale factor that is inversely proportional to vocal tract length (Fant, 1960; Lee et al., 1999). But perhaps most intriguingly, several acoustic studies report that differences between males and females are present as early as at four years of age (in F3 (Perry, Ohde, & Ashmead, 2001)), and become more apparent by age seven, where young boys are found to consistently speak with lower formants than girls, despite no significant vocal tract length dimorphism before puberty (Fitch & Giedd, 1999; Vorperian et al., 2011). These inconsistencies between acoustic and anatomical data indicate that gender-related dimorphisms in formant values cannot be explained solely by variation in VTL. While some of these differences may be related to observed localised differences between males and females in growth trend and rate for select vocal tract structures, as well as gender differences in other anatomical dimensions (e.g. vocal tract volume), the underlying causes for the mismatch between acoustic and anatomic dimorphism, especially pre-puberty, remain largely unknown (Vorperian et al, 2011).

Concomitantly, acoustic theory of voice production (Fant, 1960) and articulatory simulations (Boë & Ménard, 2000; Ménard, Schwartz, Boë & Aubin, 2007) indicate that VTL can be dynamically influenced by articulatory behaviours. Vocal tract shape and relative length of vocal tract sections can be changed voluntarily by varying the amount of mouth opening (e.g. dropping the jaw), and by movements of the body or blade of the tongue, thus affecting the relative position of the formants which generate the phonetic diversity of human speech (vowels and consonants). Of key relevance to the reported gender differences in overall formant values and spacing, is the observation that all formant frequencies can be lowered (thus narrowing ΔF) or raised (thus widening ΔF) by front-end (lip) and back-end (larynx) modifications: lip rounding and / or larynx lowering will both lengthen the vocal tract, while lip spreading and / or larynx raising will shorten its length (Hoole & Kroos, 1998; Titze, 1994). Indeed, the overall pre-pubertal dimorphism in formant frequency values has been suggested to be a consequence of these behaviours: for example, boys may protrude their lips more than girls, thus lengthening their tract and, in turn, globally lowering their formants (Lee & Iverson, 2009; Sachs et al., 1973; Whiteside, 2001). Crucially, however, acoustic research has so far focused on gender differences in individual formants, which are dependent of vocal tract shape as well as length (different formant values are required to achieve different target vowels) and therefore cannot reliably predict global adjustments in vocal tract length made by speakers. On the other hand, as a statistical measure encompassing all formant information, the spacing between formants (ΔF) is less sensitive to deviations in a single formant and thus provides a better estimate of vocal tract length achieved during utterance production (apparent Vocal Tract Length or aVTL (Fitch, 1997; Reby & McComb, 2003)).

One of the key predictions of the "gender code" hypothesis is that individuals alter the frequency components of their voice (F0 and formant parameters) by adjusting the rate of vibration of their vocal folds and by changing the apparent length of their vocal tract in order to modulate their gender and related attributes. As a preliminary investigation of this hypothesis, we propose to contrast the sex differences in the key acoustic components of male and female voices with sex differences in the morphology and dimensions of their vocal apparatus. Our aim is to highlight the extent to which acoustic differences cannot be solely explained by underlying anatomical differences. Here, we chose to focus on the comparison between vocal tract length and formant spacing, due to the availability of acoustic data (Fi values) and anatomical data (at rest MRI measurements of vocal tract length) across development. More specifically, the present study re-examines individual, vowel-specific formant frequency differences between males and females throughout development in terms of global estimates in vocal tract length achieved by speakers during utterances (Apparent Vocal Tract – aVTL, which is measured in cm) and compares them to measurements of resting anatomical vocal tract length (MRI-VTL, also measured in cm). We highlight patterns that emerge from this re-examination with reference to age and sex-linked anatomical and articulatory research, and discuss possible implications in disentangling anatomical, static measures of vocal tract length from its dynamic, behavioural variation.

Method

Resting Anatomical Vocal Tract Length (MRI-VTL)

Anatomical measurements of overall vocal tract length (MRI-VTL) come from an MRI study conducted by Fitch and Giedd (1999) on vocal tract morphology of 129 children and adults, aged two to 36. While Fitch and Giedd (1999) report mean VTL (in mm) for male and female participants grouped by age (e.g. age five to six, seven to eight), here we report VTL measurements for each individual participant (which were kindly released by the authors upon request). MRI-VTL measurements were taken by first asking subjects to lie motionless while being scanned, and to breathe quietly in order to avoid differences in measurements due to articulatory behaviour. The authors then measured vocal tract lengths of each participant as the curvilinear distance along the midline of the tract from the glottis to the intersection with a plane touching the upper and lower external borders of the lips.

Apparent Vocal Tract Length (aVTL)

The apparent Vocal Tract Length (aVTL) was measured from mean formant data of selected vowels (cardinal vowels bead /IY/, bat /AE/, pot /AA/, boot /UY/ and central vowel bet /EH/) reported in Lee and colleagues (1999). This study is the most extensive acoustic investigation to date on fundamental and formant frequency values of American-English vowels (participants were 436 children aged five to 18 years, and 56 adults aged 25 to 50 years). The apparent Vocal Tract Length was estimated from the reported mean frequency values of the first three formants (F_1, F_2, F_3) for all vowels combined and also for each vowel individually, using the regression method in Reby and McComb (2003). The method models the vocal tract as a straight uniform tube, closed at one end (the glottis) and open at the other end (the lips). According to such a model, the centre frequencies of the successive formants (F_1, F_2, ... F_i) are related to the length of the vocal tract by the equation:

$$(2)\ Fi = \frac{(2i-1)c}{4aVTL}$$

where c is the speed of sound in air (approximated as 350m/s in the human vocal tract) and aVTL is the estimated length of the vocal tract. As a consequence, the spacing between any two consecutive formants in the frequency spectrum is constant and given by:

$$(3)\Delta F = F_{i+1} - F_i = \frac{c}{2aVTL}$$

By replacing $c/2aVTL$ with ΔF in equation 2, individual formant frequencies can be related to formant spacing ΔF by:

$$(4)Fi = \frac{2i-1}{2}\Delta F$$

We can therefore estimate ΔF, the overall spacing of the formants, by seeking the best fit for equation (4) to the centre frequency of the observed formants. From equation (3) we can also deduce the apparent vocal tract length (aVTL) as:

$$(5)aVTL = \frac{c}{2(\Delta F)}$$

The Apparent Vocal Tract (aVTL) of speakers is therefore the reciprocal correlate to formant spacing. Both measurements reflect an approximation of the actual vocal tract lengths achieved by the speaker during phonation: overall, the spacing between the resonant frequencies decreases as vocal tract length increases, and increases as vocal tract length decreases. However, as aVTL is measured in cm, it is more readily comparable to anatomical measurements of VTL than ΔF (which is expressed in Hz).

Results

Sex Differences in MRI-VTL

Figure 3.1.1 illustrates individual males' and females' MRI-VTLs (dots) as well as mean MRI-VTLs for the two sexes (bold lines) throughout development. Fitch and Giedd (1999) report no significant differences in MRI-VTL between males and females before age 15 (end of male puberty), when males' vocal tracts are on average 1.25cm longer than females (15–16 yr old males' MRI-VTL: 15.40cm, 15–16 females' MRI-VTL: 14.15cm). Correspondingly, confidence intervals, represented with lighter blue (male) and green (female) lines in Figure 3.1.1, become fully distinguishable by age 15, increasing their distance from that age onwards. Fitch and Giedd (1999) show that the male vocal tract continues to go through maturational changes after puberty, and by early adulthood (19–25 years), males' VT is on average 1.48cm longer than in females (males' VTL: 16.12cm, females' VTL: 14.64cm).

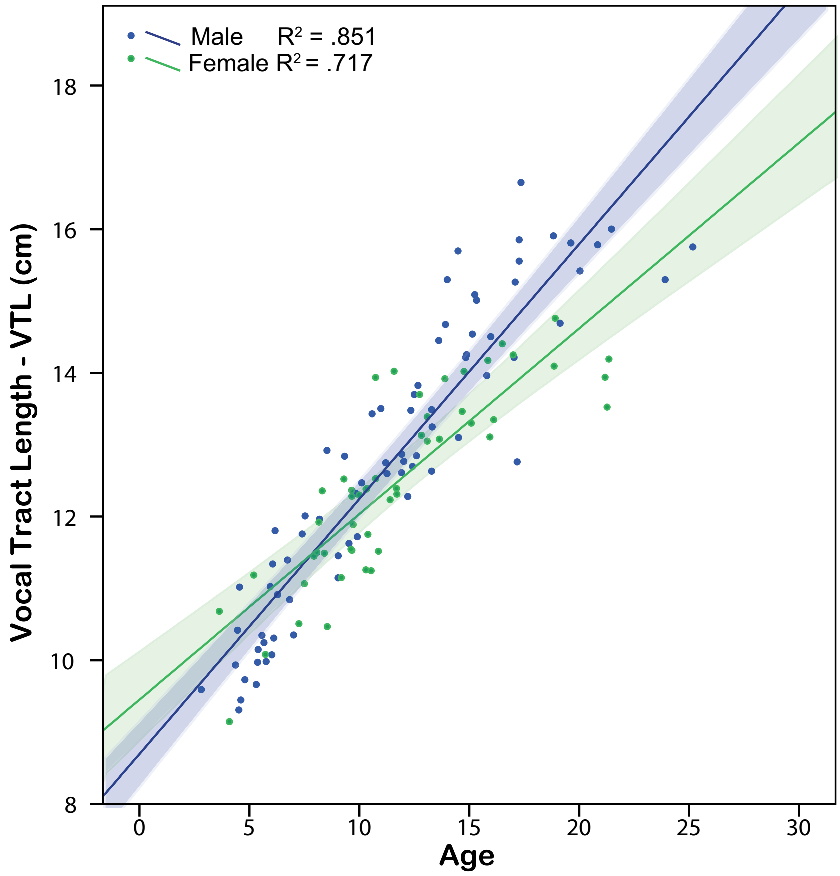

Figure 3.1.1. Individual anatomical vocal tract length (MRI-VTLs) of male (blue dots) and female (green dots) speakers, aged two to 25, with age as a continuous variable. Regression lines show estimated means in the two populations (bold lines), while confidence intervals are represented by lighter (for males) and green (for females) lines. Sex differences in overall vocal tract lengths are not significant until the end of male puberty, as shown by the confidence intervals for the two sexes largely overlapping before age 15.

Sex Differences in aVTL

Figure 3.1.2 illustrates the emergence of relatively small (about 0.5cm) between-sex differences in aVTL prior to puberty, reflecting sex differences in individual formant values of boys and girls. The dimorphism in aVTL becomes more marked (about 2cm) during the pubertal period (between 11 and 15 years of age): males' aVTL rapidly increases, while females' aVTL follows a more gradual increase, resulting in males speaking with about 2cm longer aVTLs than females overall (except for a localised dip at age 14, due to males displaying an increase in formant frequencies at that age). Sex differences in aVTL are maintained throughout the post-pubertal period, with a tendency to increase by a further 0.5cm in young adulthood, mainly due to a slight increase in males' aVTL values (while females' aVTL remains more or less stable after age 14). By the end of the observed period (25 to 50 years (Lee et al.,1999)), adult females' aVTL averages around 14.63cm, compared to 17.16cm in adult males.

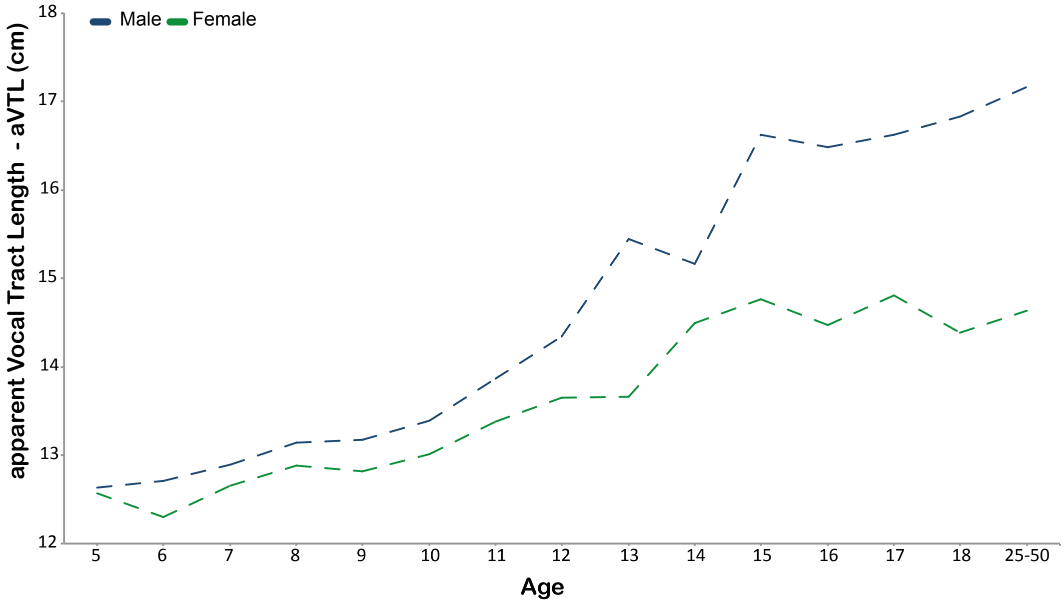

Figure 3.1.2. apparent Vocal Tract Lengths (aVTLs) averaged across vowels for males (blue line) and females (green line), aged five to 18, and 25 to 50 (together).

Figure 3.1.3 depicts aVTLs of males and females for each individual vowel under study. In line with Figure 3.1.2, overall males speak with longer aVTL than females across all vowels throughout development, with the exception of /EH/ at five and eight years of age and of /IY/ between five and eight years of age, when males spoke with shorter aVTL than females. Overall, vowels /AA/ and /UY/ display the highest sex dimorphism in aVTL, reaching 1.2cm in six to seven year olds and 3.6cm in adults. The lowest sex-dimorphism was observed for vowels /EH/ (< 0.1cm during the pre-pubertal period and < 2cm from puberty) and /IY/ (< 0.1cm during the pre-pubertal period, and < 1.8cm from puberty), for which the dimorphism between the ages of five and eight was actually reversed (females having 0.15cm longer aVTLs on average than males, due to though girls having lower F1–F3 for /IY/ than boys at age five, lower F2, F3 at age six and lower F3 at age seven).

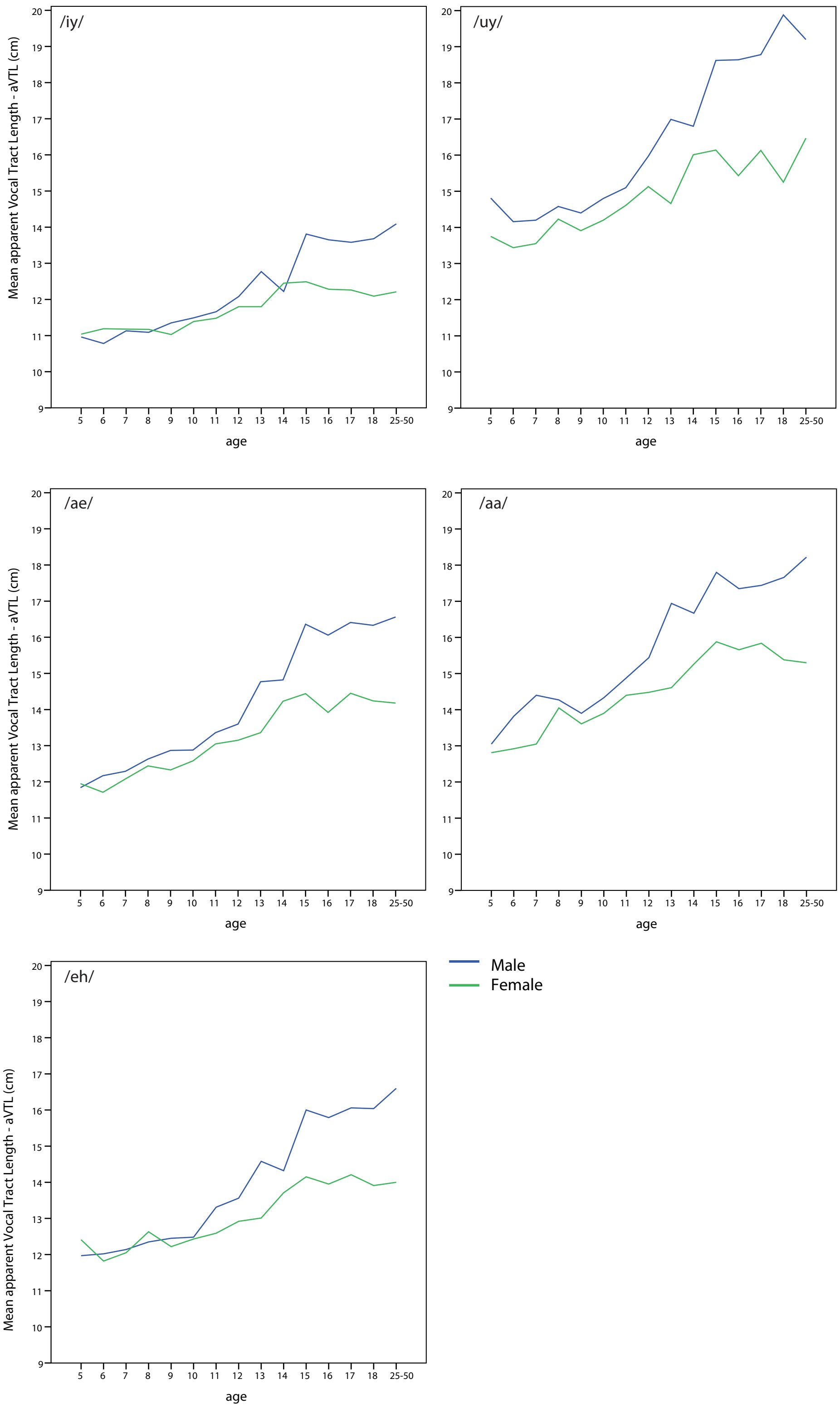

Figure 3.1.3. Apparent Vocal Tract Lengths (aVTLs) for each individual vowel as a function of sex (blue line for males, green line for females) and age (x-axis).

Discussion

The "gender" code hypothesis developed in this thesis states that speakers adjust indices to sex in the human voice in order to vary the expression of their gender and related attributes. In order to provide a clear background for the investigation of this hypothesis, this study explored whether some of the acoustic variation in overall formant spacing (ΔF), which signals gender in both pre-pubertal and adult speakers' voices, may have a behavioural component. To do this we compared ΔF's reciprocal correlate, the apparent Vocal Tract Length (aVTL), which is an estimate of anatomical vocal tract length achieved during phonation, with length measurements of anatomical vocal tract at rest. In the first instance, this comparison confirmed the relationship between the two measures, with the developmental trajectory of males' and females' aVTL closely following the general pattern of developmental sex differences in the maturation of their anatomical tracts (MRI-VTL). For example, aVTL for all vowels combined (Figure 3.1.2) as well as for individual vowels (Figure 3.1.3) gradually increase during childhood for both genders and, at the onset of male puberty, males' aVTL rapidly and steadily increases, in line with the lengthening in male MRI-VTL observed over this period, while a more gradual increase and eventual stabilisation during this period is noted in females (Figure 3.1.1). The observed close relationship between aVTL and MRI-VTL measurements confirms that apparent Vocal Tract Length in speakers' vocalisations provides accurate acoustic information about the sex of the speaker, though aVTL values are 0.5 to 1.5cm longer, as expected by the fact that VTL lengthens from its resting position during the production of most vowels (except high-front vowels e.g. /IY/ (Vorperian et al., 2011; Riordan, 1977)).

However, our results also show that sex differences in aVTLs cannot be solely explained in terms of maturational differences of the vocal tract. For example, aVTL of vowels between post puberty and adulthood vary as a function of age and sex suggesting that sex-specific developmental differences in articulation behaviours are also present. While the extent to which such behaviours are linked to the expression of voice gender remains hypothetical, it is worth noting that the highest and lowest sex differences are shown respectively for back /UY/ and front /IY/vowels, both of which allow for maximal lip and laryngeal movements to be performed while keeping the articulatory targets (See Appendix for further details on articulatory models for vowels).

For example, males' rapid increase in aVTLs for /UY/ from puberty to adulthood, compared to females' relative stable values, may indicate that males may drop their larynx towards the sternum, which allows them to fully exploit their longer pharynges compared to females, as well as rounding their lips more. Indeed, at least one sociolinguistic field study reports that adolescent male speakers of American English pronounce their /UY/ further back (lower formants) to signal masculine qualities such as "toughness" (Habick, 1991). In contrast, females' aVTL values for /IY/ tend to decrease after puberty, while increasing in males, suggesting that women may raise their larynx and spread their lips more than their male counterparts to shorten their aVTL and thus raise their formants. Indeed, facial research shows that women are reported to smile (thus spreading their lips) more than men (Hecht & La France, 1998), which in turn has been associated to them complying with gender roles e.g. expectations that women are more "gentle", "unthreatening" and "empathic" than men (Basow, 1992; Basow & Howe, 1980).

The strongest evidence for a behavioural dimension of voice gender arises from the comparison of anatomical data with acoustic measurements prior to puberty: our estimates of vocal tract lengths achieved during phonation (aVTLs) confirm that boys speak with longer tracts (and thus narrower ΔF) than girls, despite the absence of overall sex differences in anatomical vocal tract length at those ages (Fitch & Giedd, 1999; Vorperian et al., 2005; Vorperian et al., 2009; Vorperian et al., 2011). More recently, MRI-data have shown that pre-pubertal sex differences do exist in the growth trend, type and rate, as well as length of individual VT structures. More specifically, from the age of eight, males are reported to have longer and faster growing posterior cavity length (PCL), and shorter and slower growing nasopharyngeal length (NPhL), than females, though such differences are not reflected in significant differences of VT-V or indeed overall VTL until puberty (Vorperian et al., 2009), and are therefore unlikely to affect overall sex differences in formants before then. Pre-pubertal sex differences have also been found in the oral region, with males aged three to seven having longer VTH-H (the horizontal distance from a line tangential to the lips to the posterior pharyngeal wall) than females, although this difference then disappears between eight and 13 years of age (Vorperian et al., 2011). Longer VTH-H would produce lower values in individual formants associated to the front cavity, which may help explain why three to seven year old males have lower F2 values in low vowels (/AE/ and /AA/) and F3 in back vowel /UY/ than their female peers (see Appendix for further details on formant affiliations as a function of vowel production). However, it does not explain the acoustic dimorphism reported across vowels and formants between those ages (Lee & Iverson, 2009; Perry et al., 2001; Whiteside & Hodgson, 1999; Busby & Plant, 1995; Bennett, 1981).

Given the absence of relevant sex differences in the shape and dimension of the vocal apparatus before the puberty, pre-pubertal voice gender differences may have a behavioural component (Sachs et al, 1973; Lee et al, 1999; Whiteside et al, 2001). Intriguingly, research in the visual domain has shown that children control the expression of their own gender by imitating adults whose gender matches their own (Losin, Iacoboni, Martin, & Dapretto, 2012; Perry & Bussey, 1979; Slaby & Frey, 1975). Such imitative responses appear to be a key component in children's gender identity development and expression (Bussey & Bandura, 1999). As with other aspects of gendered behaviour, pre-pubertal children may therefore learn to speak as a "man" or as a "woman" by acquiring (consciously or unconsciously) articulatory behaviours during their development. Indeed, at least one study (Sachs, Lieverman & Erickson, 1973) presents anecdotal evidence that girls speak with spread lips compared to boys, thus raising their formants, mimicking facial expressions reported in adults (e.g. women speaking with a smile – Hecht & LaFrance, 1998).

Conclusion

Our results confirm the mismatch between anatomical dimorphism in vocal tract length and acoustic dimorphism of its resonances (especially prior to puberty), indicating that a substantial proportion of the acoustic variation in gender expression must result from speakers dynamically changing the length of their vocal tracts (thus affecting their formant frequencies). As such the present study provides initial support to the "gender code" hypothesis, which states that speakers make a conventionalised use of the existing sex dimorphism (F0, ΔF) to vary the expression of their gender and related attributes.

Future studies should explicitly investigate whether speakers are in fact able to dynamically modify the expression of their gender through the voice, and whether this ability develops with age, for example by measuring shifts in F0 and ΔF as children and adults are explicitly asked to sound more like a boy or girl, masculine or feminine. Acoustic data could be complemented with quantitative measurements of the articulatory features underlying the observed acoustic manipulations. For example, vocal tract length adjustments underlying ΔF manipulations could be related to quantitative measurements of lip movements, as well as laryngeal vertical movements through cine-magnetic resonance imaging (cine-MRI). Future work is also needed to establish how these vocal gestures contribute to the perception of gender, for example by asking listeners to characterise speakers' masculinity and femininity after listening to speakers' natural voices, as well as their masculinised and feminised versions. Finally, given that individuals' gender identity varies in social significance depending on the situation (Hogg, 1985; Doise, 1990), further research is warranted to explore the salience of the "gender code" in a variety of contexts, including speakers' desire to comply with specific gender roles and listeners' attentiveness to voice gender variation in the presence of gender stereotypical and counter-stereotypical information.

References

Basow, S. A. (Ed.). (1992). *Gender: Stereotypes and roles* (3rd Ed). Belmont, CA, US: Thomson Brooks/Cole Publishing Co.

Basow, S. A., & Howe, K. G. (1980). Role-model influence: Effects of sex and sex-role attitude in college students. *Psychology of Women Quarterly*, *4*(4), 558–572.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231–238.

Boë, L.-J., Ménard, L., & Maeda, S. (2000). Adaptation of control strategies during the vocal tract growth inferred from simulation studies with an articulatory model. In *5th Seminar on Speech Production* (pp. 277–280).

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*, *97*(4), 2603–2606.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676.

Doise, W. (1990). Social biases in categorization processes. *Advances in Psychology*, *68*, 305–323.

Fant G (1960) *Acoustic Theory of Speech Production.* The Hague: Mouton & Co.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology*, *6*(4), 219–225.

Gilbert, H. R., Robb, M. P., & Chen, Y. (1997). Formant frequency development: 15 to 36 months. *Journal of Voice*, *11*(3), 260–266.

Habick, T. (1991). Burnouts versus rednecks: effects of group membership on the phonemic system. *New Ways of Analyzing Sound Change*, *185*, 212.

Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, *24*, 1332–1342.

Hogg, M. A. (1985). Masculine and feminine speech in dyads and groups: A study of speech style and gender salience. *Journal of Language and Social Psychology*, *4*(2), 99–112.

Hoole, P., & Kroos, C. (1998). Control of larynx height in vowel production. In *ICSLP*.

Lee, S., & Iverson, G. K. (2009). Vowel development in English and Korean: Similarities and differences in linguistic and non-linguistic factors. *Speech Communication*, *51*(8), 684–694.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, *46*(2), 117–128.

Losin, E. A. R., Iacoboni, M., Martin, A., & Dapretto, M. (2012). Own-gender imitation activates the brain's reward circuitry. *Social Cognitive and Affective Neuroscience*, *7*(7), 804–810.

Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, *35*(1), 1–19.

Perry, D. G., & Bussey, K. (1979). The social learning theory of sex differences: Imitation is alive and well. *Journal of Personality and Social Psychology*, *37*(10), 1699.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance; Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1208.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America*, *62*(4), 998–1002.

Robb, M. P., Chen, Y., & Gilbert, H. R. (1997). Developmental aspects of formant frequency and bandwidth in infants and toddlers. *Folia Phoniatrica et Logopaedica*, *49*(2), 88–95.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 849–856.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results. *International Journal of Pediatric Otorhinolaryngology*, *49*(3), 197–206.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, *117*, 338.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging studya). *The Journal of the Acoustical Society of America*, *125*(3), 1666–1678.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research*, *54*(4), 995–1010.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, *110*, 464. doi:10.1121/1.1379087

Whiteside, S. P., & Hodgson, C. (1999). Acoustic characteristics in 6-10-year-old children's voices: some preliminary findings. *Logopedics Phonatrics Vocology*, *24*(1), 6–13.

Appendix - Articulatory Models

Because the shape of the human vocal tract resembles a tube, its acoustic properties can be estimated by modelling the tract as a tube close at one end (glottis) and open at the other (mouth), commonly defined as a “quarter-wave” resonator. For this tube, resonant frequencies (formants) are given by:

$Fi = (2i- 1)c/ 4L$

where *c* is the speed of sound in air (approximated as 350m/s in the human vocal tract) and *L* is the length of the vocal tract. For example, a vocal tract that is 17.5cm long (typical male value) will have F1 of 500Hz, F2 of 1500Hz, and F3 of 2500Hz: the resonances tend to be equidistant, and higher when the tube is shorter.

This model has proved accurate in estimating the formant frequencies of central vowels (e.g. /EH/), which are produced with no constrictions (Figure 3.1.4). Moreover, as we have seen, this model can be used to determine the average formant spacing between formants (ΔF), and its acoustic correlate Apparent Vocal Tract Length (aVTL), as averaging formants in connected speech leads to central vowel values. In turn, aVTL is a good estimate (aVTL) of “speaking” VTL, the anatomical VTL achieved during utterances (as opposed to “resting” VTL, anatomical VTL achieved during quiet breathing).

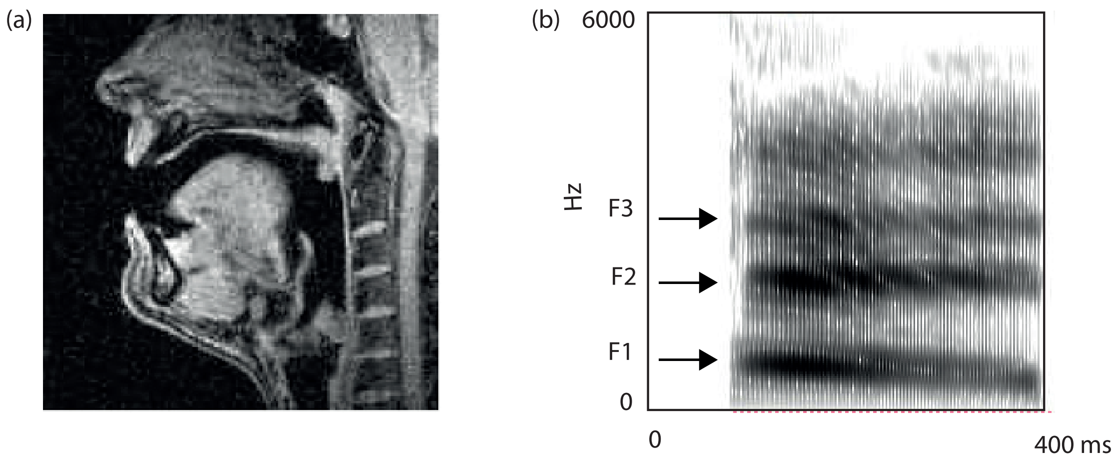

Figure 3.1.4. MRI image (a) and corresponding spectrogram (b) for vowel /EH/. This vowel is produced without vocal tract constrictions, and thus the vocal tract can be approximated to a uniform tube extending from the larynx (closed end) to the lips (open end). The spectrogram was taken from recording myself uttering the vowel /EH/ (F1:721Hz; F2:1835Hz; F3: 2696Hz).

However, the production of other vowels involve constrictions in the oral tract, affecting individual formant values and therefore requiring more complex tube models than the simple quarter-wave resonator described above. Each tube in turn will act as a filter (as explained above) and thus resulting sound will have formants that are a combination of the resonant frequencies from each tube individually.

Low Vowels

In low vowels like /AE/ or /AA/ the tongue is lowered in the front of the mouth, while bunching up at the back, thus constricting the throat. This results in a pharyngeal cavity that is relatively narrower to the oral cavity. Moreover, for a front vowel like/AE/, the length of the front cavity is short and the back cavity is long, while the reverse is true for /AA/. The tongue constriction effectively creates two tubes, each behaving like a quarter-wave resonator with its own resonant frequencies. The lowest resonances of the whole multi-tube systems are the lowest resonances of the individual tubes in the system. Thus, F1 and F3 are affiliated to the pharyngeal cavity, while F2 is affiliated to the oral cavity, as shown in the nomograms in Figure 3.1.5.

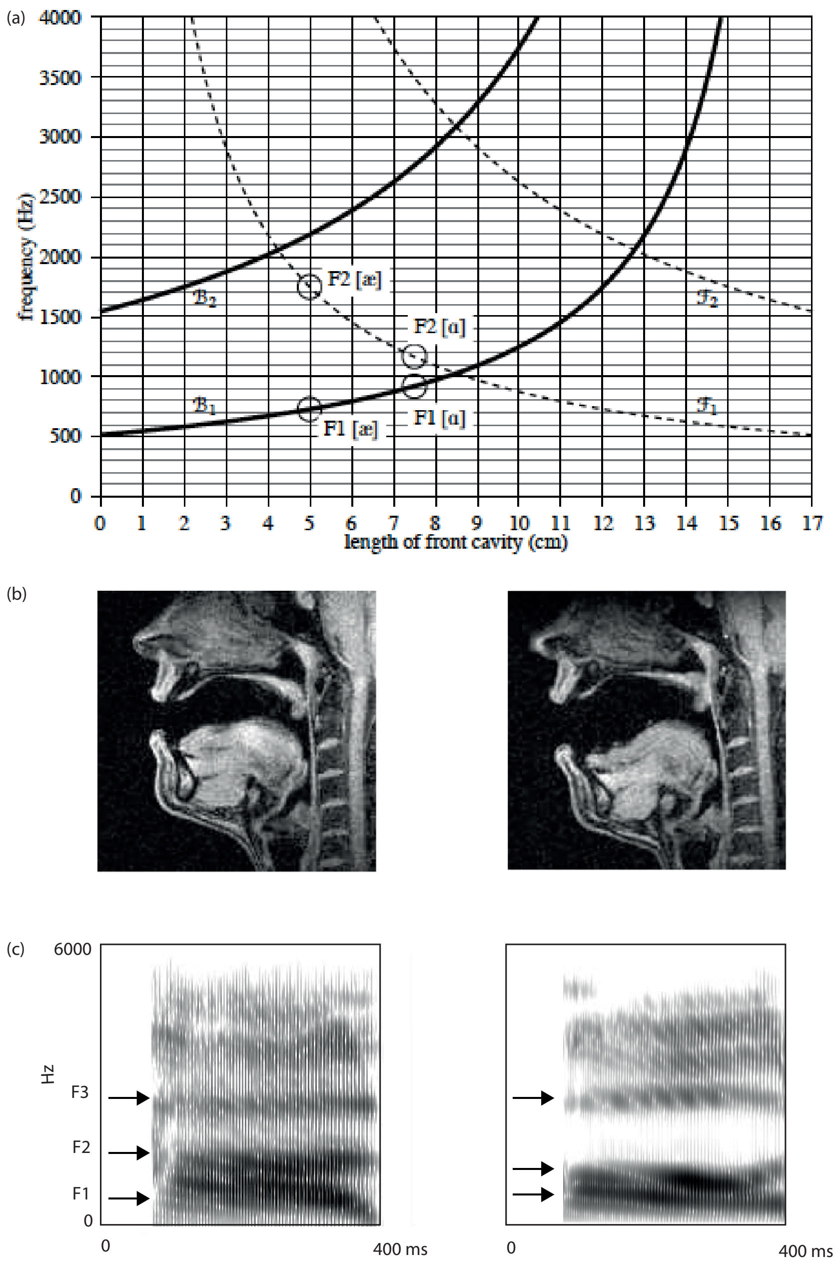

Figure 3.1.5. Nomograms (a), MRI images (b) and corresponding spectrograms (c) for vowels /AE/ on the left, and /AA/ on the right. These vowels are produced with the tongue away from the roof of the mouth, bunching up at the back. Relative to /AA/, in /AE/ the tongue is pushed forward and more lowered. My formant values for /AE/ were: F1: 728Hz; F2:1507Hz; F3: 2829Hz, and for /AA/: F1:768Hz; F2:1006Hz; F3:2891Hz.

High Vowels

In high vowels like /IY/, the tongue bunches up creating a constriction in the palatal region of the oral cavity, leaving both pharyngeal and oral cavities relatively wide (Figure 3.1.6b). This results in a three-tube model, with a small tube corresponding to the constriction itself. Thus, the back cavity is modelled as a tube closed at both ends, with a thin tube (constriction) attached to one end. The resulting structure is called a Helmholtz resonator. A Helmholtz resonator has its own resonant frequency independent of its component tubes and reflecting the relative volumes between them. For models of the mouth, the formula can be simplified as follows:

$$Fh = \frac{\alpha s}{\sqrt{l1l2}}$$

where $\ell 1$ and $\ell 2$ are the lengths of the two tubes, and α is a parameter that varies according to the openness of the constriction, between between 0 and 0.16 (e.g. the constriction is more open, α is larger, so Fh is higher). Thus form ant values for /IY/ derive from one very low resonance (F1) from the Helmholtz resonator, one high resonance from the pharyngeal tube closed at both ends (F2), and one from the oral tube closed at one end, open at the other (F3). Nomogram, MRI image and spectrogram for vowel “iy” are shown in Figure 3.1.6.

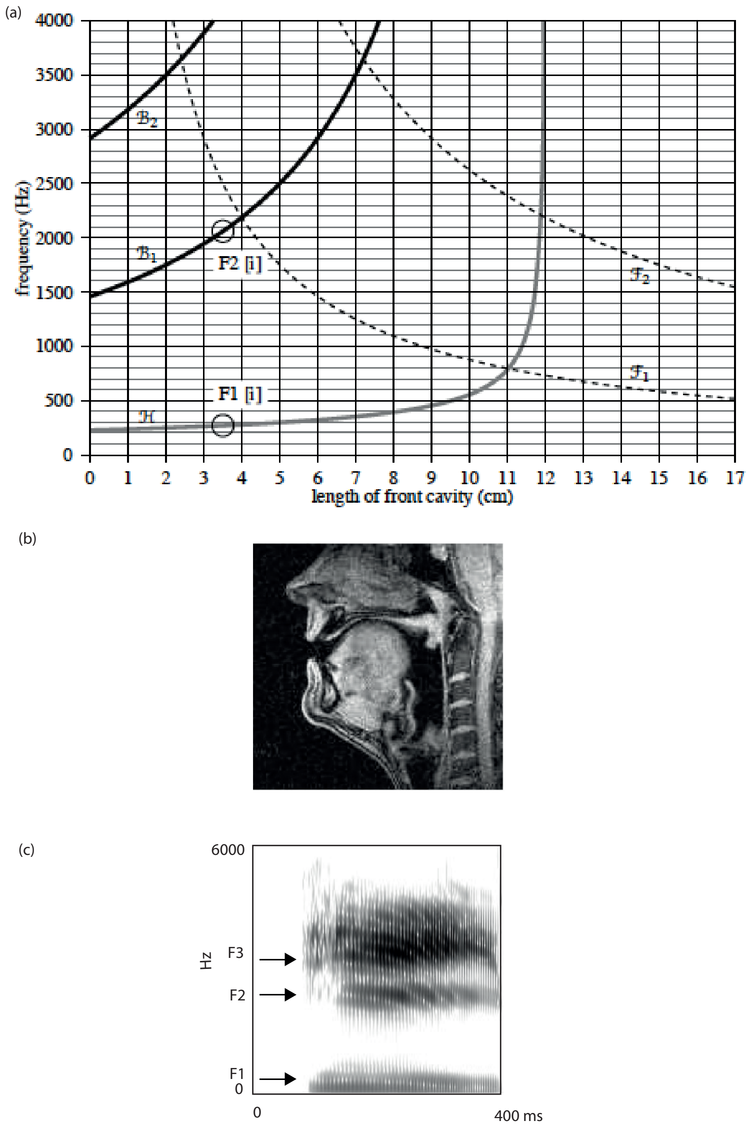

Figure 3.1.6. Nomogram (a), MRI image (b) and corresponding spectrogram (c) for vowel /IY/. This vowel is produced with spread lips, while the tongue body is raised and pushed forward, creating a palatal constriction. My formant values for /IY/ were: F1: 317Hz; F2: 2521Hz; F3: 3681Hz.

Relative to /IY/, the tongue constriction in /UY/ moves from the front to the back (pharyngeal constriction) with almost the same height (Figure 3.1.7b). However, the lips are also rounded, turning the front cavity into a fully closed tube, with a small tube corresponding to the lip constriction itself. This creates a second Helmholtz resonator in the front of the mouth, resulting in a four-tube model. In this model, F1 is affiliated to the Helmholtz resonator of the back cavity, F2 to the Helmholtz resonator of the front cavity and F3 to the front cavity itself. Nomogram, MRI image and spectrogram for this vowel are shown in Figure 3.1.7.

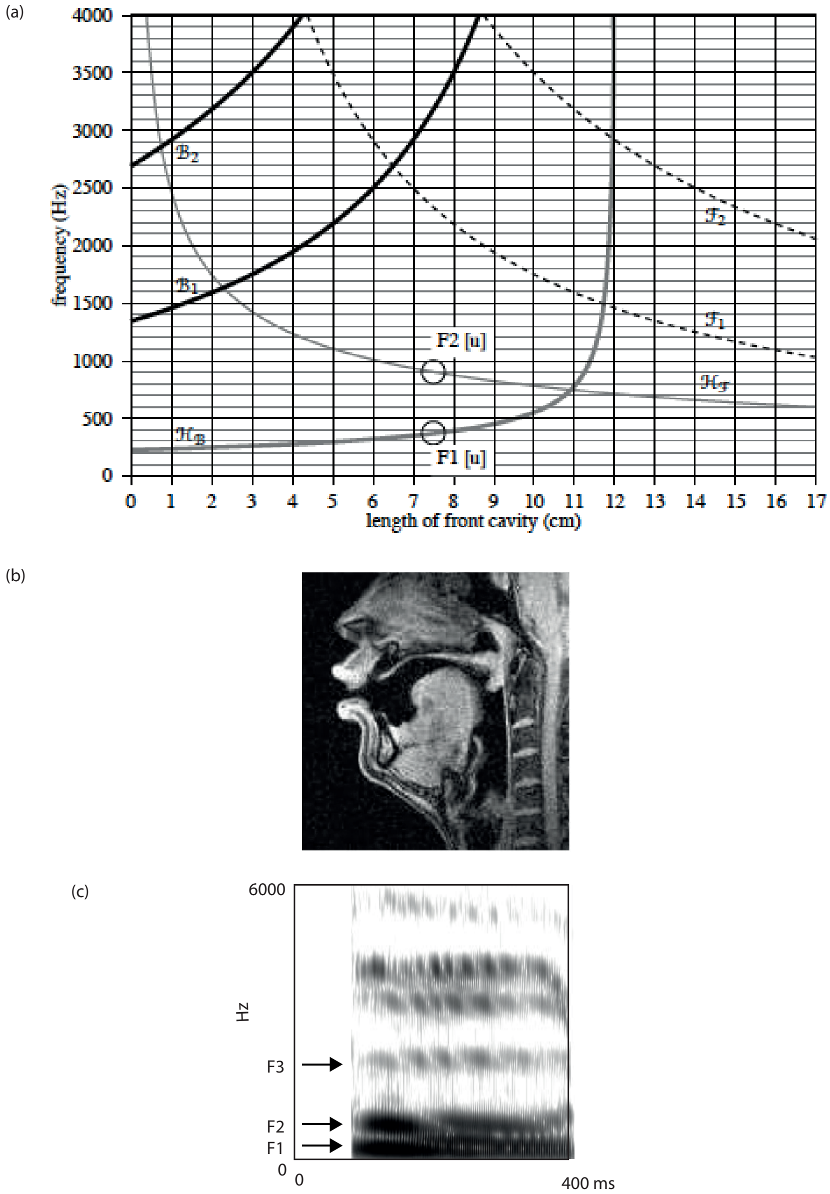

Figure 3.1.7. Nomogram (a), MRI image (b) and corresponding spectrogram (c) for vowel /UY/. This vowel is produced with rounded lips, creating a lip constriction, and the tongue body is raised and pulled back, creating a pharyngeal constriction. My formant values for this vowel were: F1: 247Hz, F2: 617Hz; F3: 2213Hz.

Note: All MRI images were adapted from *PALS1004 Introduction to Speech Science.* (n.d.). Retrieved 28 March 2014, from http://www.phon.ucl.ac.uk/. Copyright 2014 by *UCL*.

Study 2:
Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children's Voices

Note. Study published as: Cartei, V., & Reby, D. (2013). Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children's Voices. *PLoS ONE*, *8*(12), e81022.

Abstract

It is usually possible to identify the sex of a pre-pubertal child from their voice, despite the absence of sex differences in fundamental frequency at these ages. While it has been suggested that the overall spacing between formants (formant frequency spacing (ΔF)) is a key component of the expression and perception of sex in children's voices, the effect of its continuous variation on sex and gender attribution has not yet been investigated.

In the present study we manipulated voice ΔF of eight year olds (two boys and two girls) along continua covering the observed variation of this parameter in pre-pubertal voices, and assessed the effect of this variation on adult ratings of speakers' sex and gender in two separate experiments. In the first experiment (sex identification) adults were asked to categorise the voice as either male or female. The resulting identification function exhibited a gradual slope from male to female voice categories. In the second experiment (gender rating), adults rated the voices on a continuum from "masculine boy" to "feminine girl", gradually decreasing their masculinity ratings as ΔF increased.

These results indicate that the role of ΔF in voice gender perception, which has been reported in adult voices, extends to pre-pubertal children's voices: variation in ΔF not only affects the perceived sex, but also the perceived masculinity or femininity of the speaker. We discuss the implications of these observations for the expression and perception of gender in children's voices given the absence of anatomical dimorphism in overall vocal tract length before puberty.

Introduction

Adults can discriminate the sex of adult (Hillenbrand & Clark, 2009) and of child (Sachs, Lieberman & Erickson, 1973; Perry, Ohde & Ashmead, 2001) speakers by listening to their voice only. Sex identification in adult voices is substantially determined by acoustic differences in fundamental frequency (F0) and in the overall pattern of formant frequencies (ΔF, or formant spacing), which in turn reflect anatomical dimorphisms in the vocal apparatus between the two sexes. During male puberty, the testosterone-related growth of the laryngeal cartilages (Dabbs & Mallinger, 1999; Hollien, Green & Massey, 1994; Rendall, Kollias, Ney & Lloyd, 2005) and the associated lengthening and stiffening of the vocal folds (Hirano, Kurita & Nakashima, 1981) cause men's F0 to drop by almost 50% compared to women's (men's F0: 120Hz; women's: 200Hz (Titze, 1994)), conferring men their characteristically lower-pitched voices. Moreover, the testosterone-induced differential body height, with men being on average 7% taller than women (Gaulin & Boster, 1985), coupled with the male-specific secondary descent of the larynx (Fitch & Giedd, 1999), result in men having longer vocal tracts and thus narrower ΔF (15–20% (Fant, 1966; Fitch & Giedd, 1999; Goldstein, 1980)) than women, conferring a disproportionately more baritone quality to the male voice (Fitch & Giedd, 1999).

The voices of pre-pubertal children are also acoustically and perceptually different, and perceptual studies show that adults are able to correctly identify gender from the voice in children as young as four (Perry et al., 2001). Several acoustic investigations have shown that, while children of both genders speak with similar F0s (Busby & Plant, 1995; Lee, Potamianos & Naryanan, 1999; Sussman & Sapienza, 1994; but also see Tussey, Canonaco, Lynch, Oss, 2011) boys speak with lower formants and consequently narrower ΔF than girls (Bennett, 1981; Busby & Plant, 1995; Lee et al., 1999; Perry et al., 2001; Sachs et al., 1973; Whiteside & Hodgson, 2000) despite the absence of overall differences in vocal tract length between the two sexes before puberty (Fitch & Giedd, 1999; Vorperian & Kent, 2007; Vorperian et al., 2009; Vorperian et al., 2011). This dimorphism has led to the suggestion that pre-pubertal sex differences in ΔF have a behavioural basis (for example boys may round their lips or lower their larynx when they speak to lengthen their vocal tracts (Sachs et al., 1973)).

Taken together, these studies indicate that the between-sex dimorphism in the voice frequency characteristics (ΔF only in children and both ΔF and F0 in adults) is perceptually relevant to categorize the sex of speakers. Moreover, at least in adult voices, between-speaker variation in these parameters appears to also influence the perception of gender, a term which encompasses the biological and social attributes which a given society deems typical of either male (masculine attributes) or female (feminine attributes) sex (Jackson, 1998). For example, listeners consistently rate adult voices with naturally or artificially lower F0, lower ΔF, or both, as belonging to more masculine individuals than their raised versions (Pisanski & Rendall, 2011; Pisanski, Mishra & Rendall, 2012). While variation in F0 and ΔF, which are both sexually dimorphic in adult voices, has been shown to influence listeners' attributions of adults' sex and gender characteristics, to our knowledge the effect of naturalistic variation in ΔF on sex and gender attributions has not been investigated in children's voices, despite the fact that this trait is sexually dimorphic.

Here we investigate whether small increments of ΔF in children's voices affect sex (male, female), as well as gender (masculine, feminine) attributions by adult listeners. In the first experiment (sex identification) we resynthesise ΔF along gender continua within the observed natural variation of this parameter and ask listeners to identify the sex of the speakers. We expect the identification function to be characterised by a gradual change from the male to the female category. In the second experiment (gender rating), we ask listeners to rate each voice stimulus on a scale that combines sex and gender information (from "masculine boy" to "feminine girl"). We expect that small, consecutive increments in ΔF will elicit a gradual increase in listeners' ratings from "masculine boy" to "feminine girl".

Materials and Methods

Ethics Statement

Written consent from children's guardians as well as verbal consent from children were obtained prior to the recording of the voice stimuli. All adult subjects taking part in the psychoacoustic experiments gave written informed consent. Both procedures (voice recording and psychoacoustic experiments) were reviewed and approved by the Ethics Committee of the University of Sussex (authorization codes: DRVC0709 and DRVC0711).

Subjects

252 second-year Psychology students (74 males, 178 females) from Sussex University took part in the psychoacoustic experiments (as part of their practical coursework in a Cognitive Psychology level two module). All subjects were fluent English speakers.

Stimuli

Speech utterances were recorded using a *Shure SM94* microphone and a *Tascam DR07mkII* handheld recorder at a primary school in Sussex, as part of a previous study of gender expression in children's speech. During these recordings, two girls and two boys aged eight were asked to read out seven short words (“bed”, boot”, ”book”, “box”, “duck”, “hat”, “pig”). The recorded single-syllable words were individually standardized to 65 dB and concatenated prior to acoustic analysis and resynthesis.

Acoustic analyses

We extracted F0 and formant frequencies using PRAAT v.5.1.19 freeware (Boersma, 2001). F0 was extracted using the command ‘to Pitch’, with analysis parameters set to: time-step 0.01s; pitch floor, 60Hz; pitch ceiling, 500Hz. The frequency values of the first three formants (F_1, F_2, F_3) were extracted using linear predictive coding (LPC) via the ‘LPC: To Formants (Burg)’ command, with analysis parameters set to: maximum number of formants, 5; maximum formant frequencies, 6000–6600Hz; window of analysis, 0.025s. Formant spacing ((1) $\Delta F = F_{i+1} - F_i$) was derived from F_1-F_3 values, by modelling the vocal tract as a uniform tube closed at the glottis and open at the mouth (Cartei, Cowles & Reby, 2012; Reby & McComb, 2003). Under such model, F_i are expressed as:

(2) $$F_i = \frac{(2i-1)c}{4VTL}$$

Where *i* is the formant number, *c* is the speed of sound in a mammal vocal tract (35,000 cm/s), VTL is the vocal tract length (in cm) and F_i is the frequency (in Hz) of *i*th formant. From (1) and (2), it follows that $\Delta F = F_{i+1} - F_i = c/2VTL$ (3). By replacing

$c/2$VTL with ΔF in equation (2), ΔF can be derived as the slope of a regression model with the observed F_i values (y-axis) plotted against the expected formant positions:

(4) $$F_i = \frac{(2i-1)}{2}\Delta F$$

and the apparent vocal tract length (aVTL), as its inverse acoustic correlate measured in cm (aVTL = $c/2\Delta$F). Therefore the longer the vocal tract, the lower the formant frequencies, and the narrower their overall frequency spacing. All extracted and derived acoustic values are reported in Table 3.2.1.

Table 3.2.1

Acoustic variables (F0, F_i, ΔF in Hz) and apparent Vocal Tract Length (aVTL in cm) characterising the four exemplars (measured on concatenated strings of CVC words)

Exemplars	F0	F_1	F_2	F_3	ΔF	aVTL
Girl 1	237	921	2125	3381	1383	12.7
Girl 2	304	859	2099	3370	1372	12.8
Boy 1	237	786	1933	3175	1283	13.6
Boy 2	262	768	2015	3194	1302	13.4

Note. Average ΔF was 1377 Hz (aVTL 12.7 cm) for the two girl exemplars and 1293 Hz (aVTL 13.5 cm) for the two boy exemplars.

Re-synthesis

Following acoustic analysis, the stimuli were resynthesised using the "change gender" command in PRAAT. This command uses PSOLA, a resynthesis algorithm that allows the independent manipulation of formant frequency spacing (ΔF), mean fundamental frequency (F0), F0 variation and signal duration while keeping the values of all the other acoustic parameters (amplitude, noisiness etc.) unchanged. The mean fundamental frequencies were all standardised to 260 Hz (the average F0 measured in our sample). In order to remove possible intonation cues to gender, F0 variation was flattened by adjusting F0 values to the mean F0 (thus making the voice monotonous). Formant values were scaled up or down in increments of 2%, mimicking equivalent variations of ΔF (and thus aVTL) in speakers' voices. An increase of 2% of formant frequencies (achieved in the 102% stimuli) equates to a 2% increase in ΔF (corresponding to a 2% shortening of the vocal tract), and is expected to feminise the voice. As formant frequencies in our sample were on average 6% lower in the boy exemplars than in the girl exemplars, just below the gender difference reported in the literature for children of similar age (9–10% (Bennett, 1981; Perry et al., 2001)) male voices were rescaled from 88% to 118%, while female voices were rescaled from 82% to 112%. The resulting continua were therefore not identical, but largely overlapping: the boys' continuum ranged from 1526Hz to 1138Hz (aVTLs from 11.5 cm to 15.5 cm), while the girls' continuum ranged from 1542 Hz to 1129Hz (aVTLs from 11.4 cm to 15.5 cm). Supplementary online material includes audio files of example stimuli for one girl and boy exemplar. The resulting continua are within the range of ΔF variation observed in pre-pubertal children, as derived from published F1–F3 values (Lee et al., 1999), with aVTLs ranging from 11.4 cm to 15.9 cm for 5–12 year old children. They are also consistent with anatomical variation reported in Fitch and Giedd (1999), where VTLs for boys and girls, measured during quiet respiration, varied from 9.7 cm at age 5 to 14.0 cm at age 12. In summary, we generated 64 audio stimuli consisting of 16 re-synthesised variants of the single-syllable word lists by the two boys and the two girls. Figure 3.2.1 shows spectrograms of the vowel **"uh"** spoken by one of the exemplars, in which the formants (dark bands of energy in the spectrogram) are shifted compared to the original signal, while signal duration, F0 and F0 variation remain unchanged.

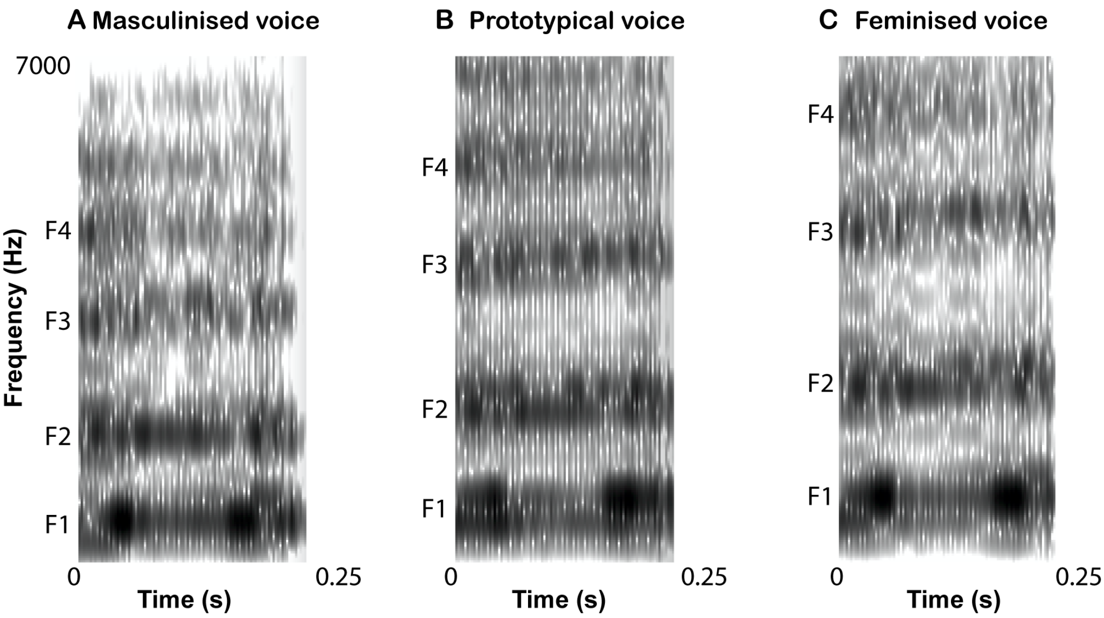

Figure 3.2.1. Spectrograms of vowel "uh" (from "book") created from girl exemplar 1. Spectrogram settings: window length = .025s, maximum number of formants, 5; maximum formant frequencies, 6000–6600 Hz. The formants (labeled F1-F4) are shifted down by 18% (A) and up by 12% (C) in comparison to the original signal (B), while all other acoustic parameters, including fundamental frequency, remain unchanged.

Procedure

Participants completed the identification experiment first. Stimuli were presented using a PRAAT Multiple Forced Choice (MFC) experiment script and for each stimulus participants were asked to decide if the speaker was male or female (the instruction was: “Please identify the sex of the speaker”) by clicking the respective button on the screen (labelled “male” or “female”). A total of different 64 stimuli (16 variants from four exemplars) were presented once in a pseudo-random order. Participants were given an opportunity to pause after each series of 32 presentations. This experiment lasted approximately 10 minutes. In the second experiment, participants were asked to rate the same 64 voice stimuli from the sex identification task (also presented in a pseudo-random order using a MFC experiment script). The instruction was: “Rate the voice of the speaker on a scale of 1 to 7” and buttons were labelled as 1= masculine boy, 2 = boy, 3 = feminine boy, 4 = neutral, 5 = masculine girl, 6 = girl, 7= feminine girl.

Statistical Analyses

Because different sets of resynthesis variants (different formant scaling factors) were used for male and female exemplars, data are analysed and reported separately by exemplar’s sex.

In order to test the effect of stimuli variant and listener sex on sex identification, we ran Generalised Linear Mixed Models (GLMM) with stimuli variant (scale), listener sex (nominal) and their interaction as fixed factors, exemplar id and subject id as random factors, and sex identification score (0 = male, 1 = female) as a binomial target variable. In order to test the effect of stimuli variant and listener sex on gender ratings we ran Linear Mixed Models (LMM) with stimuli variant (scale), listener sex (nominal) and their interactions as fixed factors, exemplar id and subject id as random factors, and gender rating as a scale outcome variable (from 1 = masculine boy to 7 = feminine girl).

Simple logistic regressions (one for boy exemplars and one for girl exemplars) were then used to illustrate the relationship between formant frequency spacing and identified sex with average score (over all participants) as the dependent variable and stimuli variant as the independent variable. Logistic models provide estimates for the slope of the category (here ‘male’ to ‘female’) transition (b1 coefficient, ranging between 0 and 1, with lower values reflecting steeper transitions) (Keating, 2004; Mullennix, Johnson, Topcu-Durgun & Farnsworth, 1995; Smits, Sereno & Jongman, 2006) and for the perceived category boundary (where 50% of stimuli are categorised a male, and 50% as female). The category boundary was computed using the formula -Ln(b0)/Ln(b1) where b0 is the constant of the logistic curve and b1 is the coefficient related to the slope (Aliaga-Garcia & Mora, 2009; Keating, 2004). Simple linear regressions with stimuli variant as the predictor variable and average gender ratings (over all the participants) as the outcome variable were used to illustrate the relationship between formant frequency spacing variant and perceived gender. All the statistical analyses were performed using SPSS v.20.0.

Results

Sex Identification Experiment

The results of the GLMM on sex identification scores of boy exemplars revealed a significant main effect of stimuli variant, $F(1, 8.060) = 2{,}696.66$, $p < .001$, while no significant main effects of listener's sex, $F(1,8.060) = 2.50$, $p = .114$, and of its interaction with stimuli variant, $F(1, 8.060) = 3.47$, $p = .063$, were found. A logistic regression (Figure 3.2.2 – black line) provided a strong statistical fit for the observed relationship between stimuli variant and average sex identification scores, $R^2 = .95$, $F(1, 14) = 240.43$, $p < .001$. The relatively shallow transition (b1 = .65) from one response category to the other indicates that the percentage of stimuli identified as female increases progressively as ΔF increases. Using this model, the estimated ‘‘male-female” boundary fell between stimulus 11 and 12 (-Ln(127.43)/(.65) = 11.25, where b0 = 127.43 and b1 = .65, corresponding to 108%–110% variants or ΔF ~ 1400Hz).

The results of the GLMM on sex identification scores of girl exemplars revealed a significant main effect of stimuli variant, $F(1, 8.060) = 1{,}869.28$, $p < .001$, while no significant main effects of listener's sex, $F(1, 8.060) = 1.99$, $p = .158$, and of its interaction with stimuli variant, $F(1, 8.060) = 2.04$, $p = .153$, were found. A logistic regression (Figure 3.2.3 – black line) provided a strong statistical fit for the observed relationship between stimuli variant and average identification scores, $R^2 = .97$, $F(1, 14) = 382.14$, $p < .001$. The relatively shallow transition (b1 = .67) from one response category to the other indicates that the percentage of stimuli identified as female increases progressively as ΔF increases. Using this model, the estimated ‘‘male-female” boundary fell between stimulus 7 and 8 (-Ln(17.37)/Ln(.67) = 7.13, where b0 = 17.37 and b1 = .67, corresponding to 94%–96% variants or ΔF ~ 1300Hz).

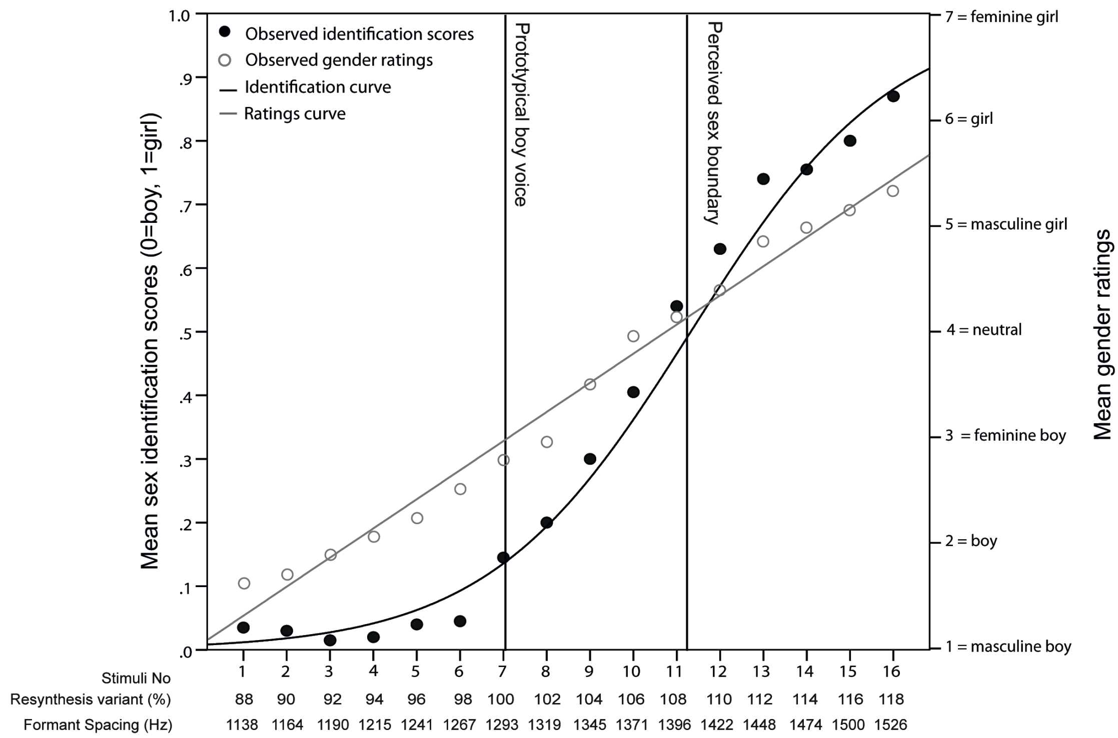

Figure 3.2.2. Identification and rating scores of boys’ voices along the gender continua. Scores were averaged across listeners on voice stimuli (numbered 1–16 on the x-axis) for the boys’ exemplars. The mean identification scores are plotted from 0=male to 1=female (left y-axis) and fitted with the logistic curve (black line). The vertical lines illustrate the location of the estimated sex boundary (where 50% of the listeners rate the stimuli as female) and the location of the prototypical boy voice stimulus (100%). The percentage of stimuli identified as female follows an S-shaped pattern along the continuum of resynthesis variants. The sex identification curve is characterised by a lower plateau for stimuli 1 to 6 (ΔFs of 1138–1267 Hz), where less than 10% of the stimuli are identified as female, indicating that stimuli variant with the lowest ΔF are mostly identified as male. The percentage of stimuli identified as female then increases gradually and linearly, and while no upper plateau is reached, average scores for stimuli 14 to 16 (ΔFs of 1474–1526 Hz) varied from 76% to 85%, indicating that boys' voices with the highest ΔF are mostly classified as female. Average gender rating scores are plotted from 1=masculine boy (or girl) to 7=feminine boy (or girl) (right y-axis) and fitted with a linear function (straight grey line). Mean gender ratings of male voices ranged from 1.78 (SE=.07) for the lowest ΔF variants to 5.36 (SE=.08) for the highest ΔF variants.

Gender Rating Experiment

The results of the LMM on gender ratings of boy exemplars revealed a significant main effect of stimuli variant, $F(15, 7781) = 692.41$, $p < .001$. No significant main effect of listener's sex, $F(1, 250) = 2.24$, $p = .136$, and of its interaction with stimuli variant, $F(1, 7781) = 1.136$, $p = .317$, were found. The results of the LMM on gender ratings of girl exemplars revealed a significant main effect of stimuli variant, $F(15, 7781) = 626.87$, $p < .001$. No significant main effect of listener's sex, $F(1, 250) = .196$, $p = .658$, and of its interaction with stimuli variant, $F(1, 7781) = .714$, $p = .773$, were found. Simple linear regressions (Figures 3.2.2 and 3.2.3 – grey straight lines) provided strong statistical fits for the observed correlation between variant number and average gender rating scores, showing that scores increased (from masculine boy to feminine girl) as formant frequency spacing increased (male exemplars: $R^2 = .99$, $F(1, 14) = 893.04$, $p < .001$, female exemplars: $R^2 = .97$, $F(1, 14) = 459.94$, $p < .001$).

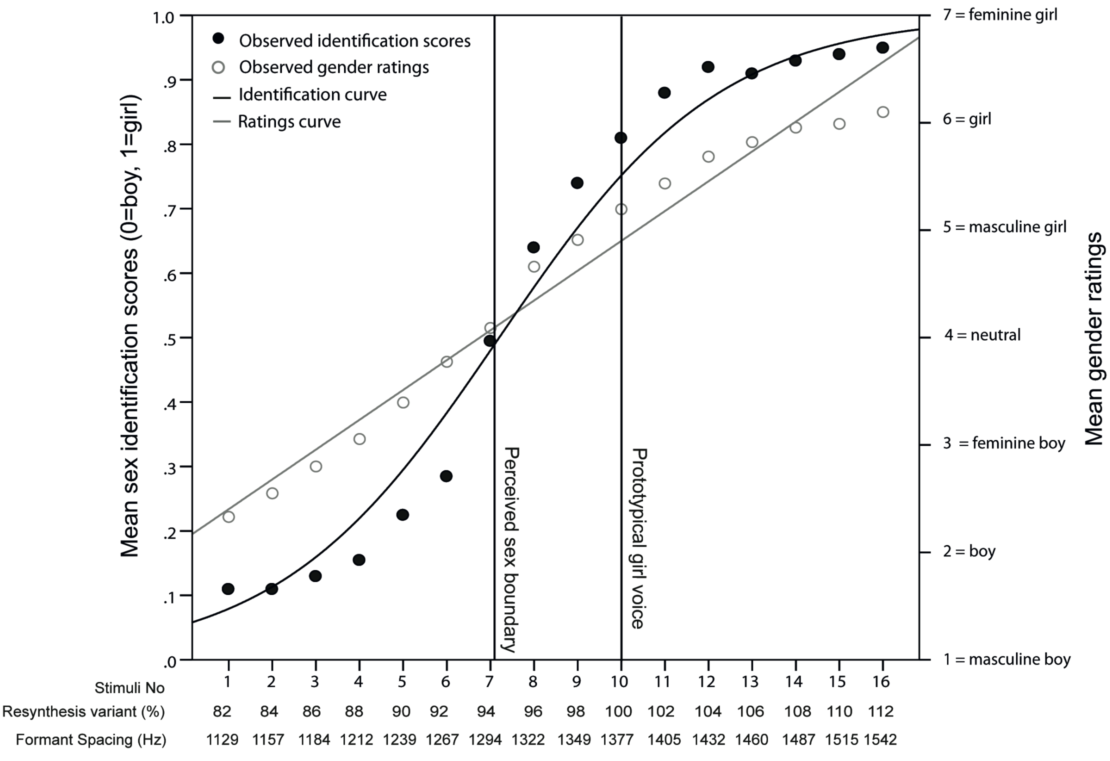

Figure 3.2.3. Identification and rating scores of girls' voices along the gender continua. Scores were averaged across listeners on voice stimuli (numbered 1–16 on the x-axis) for the girls' exemplars. The mean identification scores are plotted from 0=male to 1=female (left y-axis) and fitted with the logistic curve (black line). The vertical lines illustrate the location of the estimated sex boundary (where 50% of the listeners rate the stimuli as female) and the location of the prototypical boy voice stimulus (100%). The percentage of stimuli identified as female also follows an S-shaped pattern along the continuum of

resynthesis variants. The sex identification curve is characterised by a lower plateau for stimuli 1 to 3 (ΔFs of 1129–1184 Hz), where between 10% and 15% of the stimuli are identified as female, indicating that stimuli variant with the lowest ΔF are mostly identified as male. The percentage of stimuli identified as female then increases gradually and linearly until it reaches an upper plateau from stimuli 12 to 16 (ΔFs of 1432–1542 Hz), with average scores varying from 92% to 95% and indicating that girl voices with the highest ΔF are mostly classified as female. Average gender rating scores are plotted from 1=masculine boy (or girl) to 7=feminine boy (or girl) (right y-axis) and fitted with a linear function (straight grey line). Mean gender ratings of female voices ranged from 2.33 (SE=.02) for the lowest ΔF variants to 6.10 (SE=.06) for the highest ΔF variants.

Discussion

The results of the sex identification and gender rating experiments show that ΔF is an important cue for the perception of sex and gender in the pre-pubertal human voice, in line with the previously reported acoustic dimorphism of this parameter in pre-pubertal speakers (Lee et al., 1999; Titze, 1994; Whiteside & Hodgson, 2000; Whiteside, 2001). More specifically, the absence of a sharp boundary between the sex categories in the identification experiment, in which listeners were asked to identify the child speaker as male or female, suggests that small, sex-related acoustic variation in ΔF proportionally affects the probability of voices to be perceived as either male or female by raters. Additionally, the gradual slope in voice ratings from “masculine boy” to “feminine girl” in the second experiment shows that small linear increments in ΔF also proportionally affect listeners' attributions of speakers' gender (from masculinity to femininity). Similar results have been reported in studies of gender perception in adult voices. A study using a combination of identification and discrimination paradigms (Mullenix et al., 1995) found that variations along a male-female continuum of F0 and ΔF, the main cues to sex in adult voices, were not remapped by listeners into separate psychological (male or female) categories, indicating that the perception of voice sex was not categorical. Moreover, psychoacoustic studies have shown that both men's and women’s voices with naturally low, or artificially lowered, F0 and ΔF (or both), are rated as more masculine (Munson, 2007; Pisanski et al., 2012; Pisanski & Rendall, 2011).

In the present study, while the resynthesis continua used for boy and girl exemplars were largely overlapping (boys: 1138–1526 Hz; girls: 1129–1542 Hz) and both comprised within the range of ΔF values achievable by both genders before puberty (Fitch & Giedd, 1999; Lee et al., 1999), the effect of the rescaling of ΔF differed between boy and girl voice exemplars, suggesting that the resynthesis of this parameter was not sufficient to produce a voice systematically perceived as belonging to the opposite sex, despite the standardisation of F0 and its variation. In the sex identification experiment, the perceived sex boundary between male and female identification estimated by the logistic model is ~100 Hz higher in boy voice exemplars than in girl voice exemplars (Figure 3.2.2 – vertical lines), revealing that a greater upward shift in ΔF was required for resynthesised stimuli from the voices of the two boy exemplars to be perceived as female. The identification curve (Figure 3.2.2 – black line) for the male exemplars is also shifted downwards relative to that of the female exemplars (Figure 3.2.3 – black line), with a wider plateau at the lower (male) end of the continuum, and no plateau at the upper (female) end of the continuum. Further, the boys' rating function (Figure 3.2.2 – grey straight line) from the gender rating experiment is shifted downwards compared to girls', revealing that stimuli from boy exemplars were perceived as more masculine than those from girl exemplars. One possible explanation for the observed perceptual differences is that listeners were affected by acoustic factors other than those manipulated (ΔF) or factored out (F0 and its variation) in the present experiments. For example, Klatt & Klatt (1990) report that women are perceived to have more breathy voices than men, corresponding to increased F_1 bandwidths and decreased F_1 amplitude, while breathy voices are judged as more feminine than less-breathy voices (Van Bordel, Janssens & De Bodt, 2009), suggesting that, at least in adults, breathiness may be a contributing factor to the perception of sex and gender. The potential role of parameters such as F0, F0 variation and breathiness (Klatt & Klatt, 1990; Titze, 1994), which are sexually dimorphic in adults, but not in pre-pubertal children (Busby & Plant, 1995; Sussman & Sapienza, 1994), in the attribution of sex and gender to children's voices, is an important area for future research.

Independently from other hypothetical voice cues to sex and gender attributions of pre-pubertal children's voices, this study clearly identifies a substantial effect of ΔF variation on adults' ratings of gender in pre-pubertal speakers, with lower ΔF being consistently rated as belonging to more masculine children. ΔF variation has also been shown to affect judgements of body size and age in adult speakers, with listeners rating lower ΔF as belonging to older and larger individuals (Collins & Missing, 2003; Rendall, Vokey & Nemeth, 2007; Simmons, Peters & Rhodes, 2011; Smith, Patterson, Turner, Kawahara & Irino, 2005). These perceptual differences in turn appear to relate to actual differences in age and size of speakers (Lienard, Lacroix, Kreutzer & Leboucher, 2006; Collins & Missing, 2003; Rendall, Kollias, Ney & Lloyd, 2005). By extending the present paradigm to include age and body size ratings, future studies could investigate the perceptual linking of age-related size and gender dimensions, for example whether children that are perceived to be more masculine are also perceived to be older and bigger than their more feminine counterparts. Moreover, the use of natural (rather than re-synthesised) stimuli from children of different ages, body sizes and masculinities (e.g. as assessed by children's personal attributes questionnaires (Hall & Halberstadt, 1980)), and of raters of different ages, would help clarifying the extent to which ΔF reliably cues for these dimensions throughout the lifespan.

Our observations that baseline ΔF variation within the natural range of children's voices affects listeners' sex and gender attributions (despite the absence of a clear anatomical basis for such variation) lends further support to the hypothesis that sex and gender expression in pre-pubertal children's voices have a strong behavioural, acquired dimension (with children learning to adjust their VTL in order to sound more or less feminine/masculine). Future studies using e.g. structural cine 3D structural MRI are now needed to further test this hypothesis.

Furthermore, it has been shown that children can also spontaneously modify ΔF (and F0) when asked to sound more or less like a boy or girl (Cartei, Cowles, Banerjee & Reby, 2013), suggesting that children can also control the gender-related characteristics of their voices. The extent to which this ability affects the expression of gender in everyday speech, in line with varying gendered roles (e.g. to affiliate with same-sex peers) and contexts (e.g. when speaking to a male or female), and its perceptual relevance in gendered attributions remains to be investigated.

References

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent Research in Second Language Phonetics/phonology: Perception and Production*. Newcastle upon Tyne:

Cambridge Scholars Publishing, 2–31.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America, 69*(1), 231–238.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 19)[Computer program]. Retrieved May 1, 2010.

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences, 273*(1582), 83–89.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America, 97*(4), 2603–2606.

Cartei, V., Cowles, H. W., & Reby, D. (2012). *Spontaneous voice gender imitation abilities in adult speakers. PLoS ONE, 7*(2), e31353.

Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2013). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology, 32*(1).

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*(5), 997–1004.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences, 27*(4), 801–804.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511. doi:10.1121/1.427148

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology, 6*(4), 219–225.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children*. Massachusetts Institute of Technology.

Hall, J. A. & Halberstadt, A. G. (1980). Masculinity and femininity in children: Development of the Children's Personal Attributes Questionnaire. *Developmental Psychology, 16*, 270.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics, 71*(5), 1150–1166.

Hirano, M., Kurita, S., & Nakashima, T. (1983). Growth, development and aging of human vocal folds. *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, 22–43.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America, 96*(5), 2646–2654.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857.

Jackson, L. A. (1992). *Physical appearance and gender: Sociobiological and sociocultural perspectives*. SUNY Press.

Lee, E. A. E., & Troop-Gordon, W. (2011). Peer processes and gender role development: Changes in gender atypicality related to negative peer treatment and children's friendships. *Sex Roles, 64*(1-2), 90–102.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America, 98*(6), 3080–3095.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech, 50*(1), 125–142.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America, 109*, 2988. doi:10.1121/1.1370525

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*.

Pisanski, K. & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*, 2201.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*(3), 519–530. doi:10.1006/anbe.2003.2078

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America, 117*(2), 944–955.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Simmons, L. W., Peters, M., & Rhodes, G. (2011). Low Pitched Voices Are Perceived as Masculine and Attractive but Do They Predict Semen Quality in Men? *PLoS ONE, 6*(12), e29271.

Smith, D. R., Patterson, R.D., Turner, R., Kawahara, H., Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America, 117*, 305.

Smits, R., Sereno, J., Jongman, A. (2006). Categorization of sounds. Journal of Experimental Psychology Human Perception and Performance, 32, 733.

Sussman J.E. & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice, 8*, 145–156.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Tussey, R., Canonaco, J., Lynch, C., & Oss, E.V. (2011). Out of the Mouths of Babes: Acoustic Variation in Child Speakers. *10th Annual Celebration for Undergraduate Research and Creative Performance*.

Van Borsel, J., Janssens, J. & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: a perceptual approach. *Journal of Voice, 23*, 291–294.

Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language, and Hearing Research : JSLHR, 50*(6), 1510–1545.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging studya). *The Journal of the Acoustical Society of America, 125*(3), 1666–1678.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: an imaging study. *Journal of Speech, Language and Hearing Research, 54*(4), 995.

Whiteside, S. P., & Hodgson, C. (2000). Some acoustic characteristics in the voices of 6-to 10-year-old children and adults: a comparative sex and developmental perspective. *Logopedics Phonatrics Vocology, 25*(3), 122–132.

30. Keating, P. (2004). *Analysis of identification data*. Retrieved from http://www.linguistics.ucl.edu/faciliti/facilities/statistics.ident.htm

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America, 110*, 464.

Study 3:

What makes a voice masculine: multilevel investigation of physiological and acoustical bases of perceived masculinity

Note. Study under revision for the *Journal of Hormones and Behavior* as: Cartei, V., Bond, R. & Reby, D. (2013). What makes a voice masculine: multilevel investigation of physiological and acoustical bases of perceived masculinity.

Abstract

Men's sexually dimorphic voice contains acoustic cues to body size and hormonal status, which have been found to affect women's ratings of speaker size, masculinity and attractiveness. However, the extent to which these voice parameters mediate the relationship between speakers' fitness-related features and listener's judgments of their masculinity has not yet been investigated.

We audio-recorded 37 adult heterosexual males performing a range of speech tasks and asked 20 adult heterosexual female listeners to rate speakers' masculinity on the basis of their voices only. We then used a two-level (speaker within listener) path analysis to examine the relationships between the physical (testosterone, height), acoustic (fundamental frequency or F0, and resonances or ΔF) and perceptual dimensions (listeners' ratings) of speakers' masculinity. Overall, results revealed that male speakers who were taller and had higher salivary testosterone levels also had lower F0 and ΔF, and were in turn rated as more masculine. The relationship between testosterone and perceived masculinity was almost entirely mediated by F0, while that of height and perceived masculinity was partially mediated by both F0 and ΔF.

These observations confirm that women listeners attend to sexually dimorphic voice cues to assess the masculinity of unseen male speakers. In turn, variation in these voice features correlate with speakers' variation in stature and hormonal status, highlighting the interdependence of these physical, acoustic and perceptual dimensions.

Introduction

Male masculinity is typically associated with the expression of sexually selected morphological traits that emerge at sexual maturity (Andersson, 1994) and which are associated with individuals' hormonal and physical quality. For example, masculine facial (e.g. large jaws and pronounced brows) and bodily (e.g. broad shoulders and narrow hips) traits positively correlate with testosterone levels, health status, disease resistance, physical strength and self-reported mating success (Fink et al., 2003; Fink, et al., 2007; Gallup et al., 2007; Hönekopp et al., 2007; Penton-Voak & Chen, 2004; Prokop et al., 2013; Rantala et al., 2012; Thornhill & Gangestad, 2006). To the extent that masculinity correlates with underlying fitness, perceiving its variation is crucial when choosing a mate. Indices of masculinity in men's faces and bodies are indeed attractive to women, especially when most fertile during their menstrual cycle (Little et al., 2007; Welling et al., 2007; Zebrowitz & Rhodes, 2002) and when explicitly asked to judge for short term mating (Little et al., 2002; Rhodes et al., 2005).

Along with facial and bodily features, the human voice is a sexually dimorphic trait: compared to women, men speak at a lower fundamental frequency or F0 (lower pitch), and lower, more closely spaced formant frequencies (deeper timbre (Fitch & Giedd, 1999; Titze, 1994)). These differences are at least partly affected by hormonally induced changes occurring during male puberty. Pubertal exposure to androgens causes a 60% increase in men's vocal fold length relative to women's, and a corresponding decrease in its inverse acoustic correlate, mean F0 (Harries et al., 1998; Titze, 1994). Under the influence of androgens, pubertal males also grow 7% taller than females on average (Gaulin & Boster, 1985) and develop a further descended larynx, causing an increase in the lengthening of their vocal tract and thus a permanent drop in its inverse acoustic correlate, formant spacing or ΔF (Fitch & Giedd, 1999; Vorperian, 2009). Because of the relationship between sexually dimorphic acoustic properties and underlying biological dimorphisms, acoustic variations among adult males may provide indexical cues of fitness-related features (e.g. testosterone levels, mating success, body size), with lower frequency (more masculine) values signalling greater fitness. Indeed, men's individual mean F0 has been found to negatively correlate with circulating levels of testosterone (Dabbs & Mallinger, 1999; Evans et al., 2008; Puts et al., 2012) and higher mating success rates (Apicella et al., 2007; Hodges-Simeon et al., 2011). At least one study (Bruckert et al., 2006) has also reported a negative relationship between ΔF and testosterone, though more recent studies have failed to replicate these findings (Evans et al., 2006; Puts et al., 2012). At the same time, ΔF seems to moderately correlate with speakers' body size, and in particular men's height (Collins & Missing, 2003; Greisbach, 2007; Rendall et al., 2005; but see Van Dommelen & Moxness, 1995), with taller men speaking with lower ΔF, while there appears to be no consistent relationship between body size measures and F0 (Rendall et al., 2005; Puts et al., 2012; Van Dommelen & Moxness, 1995). If vocal frequencies signal hormonal and physical attributes of the speaker, attending to such acoustic cues may have important consequences when assessing potential mates. Indeed, psychoacoustic studies (where voice frequencies are artificially manipulated) report that pronounced sexually dimorphic (more masculine) features in men's voices positively affect women's masculinity ratings (Feinberg et al., 2005; Feinberg et al, 2006; Feinberg et al., 2008; Jones et al., 2010), as also shown for men's faces and bodies (Feinberg et al., 2008; Little et al., 2002; Little et al., 2007). However, the complex relationships among fitness-related, acoustic and perceived dimensions of males' masculinity remain under-investigated. The present study tests the hypothesis that the natural variation in sexually dimorphic voice cues (F0 and ΔF) of male speakers mediates the effects of their fitness-related characteristics (testosterone and height) on masculinity attributions made by women listeners. More specifically, we expect taller, higher-testosterone, men to speak with lower frequency values (with testosterone mainly cueing for F0 and body height for ΔF), and to receive higher masculinity ratings, than their shorter, lower-testosterone peers.

Methods

Participants

We recorded voices from 37 self-reported heterosexual men with no history of chronic diseases or hormonal abnormalities, all native speakers of British English and aged 20 to 25 (*M* = 20.6, *SD* = 1.7). None were currently suffering from any conditions that might affect their voice (e.g. colds, sore throats). Listeners were 20 undergraduate female students, aged 20 to 25 (*M* = 21.8, *SD*.= 1.5) from the University of Sussex, Brighton (UK). All women were self-reported heterosexuals, with no history of hearing impairments and with British English as their first language. All participants gave their written informed consent prior to taking part in the production and perception experiments. Approval for both procedures was granted by the School of Life Sciences Research Governance Committee (Certificates of approval: DRVC0409 and DRVC0711).

Physical Masculinity

Speakers were individually audio-recorded in a soundproofed booth at the University of Sussex. Prior to the recording of their voices, participants' body height was measured to the nearest 0.1cm using a *Seca Leicester* stadiometer, from the top of the participant's head to the soles of his feet (shoes off and feet together), with the participant standing erect and looking straight ahead. Saliva samples were taken from speakers immediately after the recordings. Participants were asked to confirm that they had not eaten, drank, chewed gum or brushed their teeth for at least 30 minutes before sampling, and were asked to rinse their mouth for 10 seconds prior to collection. Collection was performed using a *Salimetrics Oral Swab* (SOS) under the front of the speakers' tongue: speakers kept the swab in their mouth for three minutes (without chewing it), and then placed it in its plastic storage tube, without touching the swab with their hands. Samples were stored in a freezer at – 20°C and sent to Salimetrics for testosterone analysis via Immunoessay (Salimetrics). All saliva collections were carried out between 9 am and 11 am, to control for the effect of diurnal variation in F0 and testosterone levels (Evans et al., 2008). Means and standard deviations for body height and salivary testosterone levels across the 37 speakers are reported in Table 3.3.1.

Table 3.3.1

Means and standard deviations for body height (cm) and salivary testosterone levels (pg/ml)

Physical measures	N	Range	Mean	SD
Height(cm)	37	170.50–190.10	180.10	4.80
Testosterone (pg/mL)	37	87.10–253.25	153.60	40.69

Voice Masculinity

Recordings of male speakers were taken in a soundproofed room using an *AKG PERCEPTION 220* microphone. Speakers were asked to read out loud the words *had, head, hid, heed, hod, hood, who'd*, followed by the Rainbow passage (Fairbanks, 1960). Next, in order to elicit spontaneous speech (rather than text read aloud) while obtaining the same phonetic data (LListerri, 1992), subjects were given a picture of a kettle, and asked to describe it for 60 seconds, ending the description by answering the question "*what is the object in front of you?*". Three types of voice stimuli were created from these recordings in order to be used in the rating phase of the study: a list of single-syllable words concatenated with an interval of 0.5s silence (isolated word stimuli), the sentence "*people look, but no-one ever finds it*" extracted from the Rainbow passage (sentence stimuli), and the statement "*the object I have in front of me is a kettle*" from the spontaneous description of a kettle (spontaneous speech stimuli). Thus, a total of 111 audio samples (37 speakers x 3 types of voice stimuli) was used in the voice ratings. Stimuli were individually standardised to 65dB prior to acoustic analysis. Fundamental frequency (F0) values and the frequency of the first four formants ($F1$–$F4$) were obtained from these stimuli, using a custom script in *PRAAT v.5.2.17* (Boersma & Weenink, 2011) for batch processing. The computed values were double checked by visual inspection of the spectrogram and analysis parameters adjusted accordingly to correct erroneous estimates. Mean fundamental frequency (F0) was calculated using the *PRAAT* autocorrelation algorithm ''to Pitch''. The analysis parameters were set as pitch floor 30 Hz and ceiling 500 Hz, time step 0.01 s. The frequencies of the first four formants were obtained using *PRAAT*'s Linear Predictive Coding ''Burg'' algorithm. The analysis parameters were set as: number of formants=5, maximum formant=5000 Hz, and dynamic range=30 dB, length of the analysis window=0.03 s. The centre frequencies for *F1–F4* of each sample were used to derive average formant spacing (ΔF), that is, the distance between any two adjacent formants ($\Delta F = F_{i+1} - F_i$), by seeking the best fit for the equation:

$$F_i = \frac{2i-1}{2}\Delta F$$

(see Cartei et al., 2012; Reby & Mcomb, 2003 for details). Mean acoustic values across speakers are reported in Table 3.3.2.

Table 3.3.2

Means and standard deviations for the acoustic parameters in the three voice stimuli (isolated words, sentence, and spontaneous speech)

Acoustic parameters	Voice stimuli					
	Isolated words (N=37)		Sentence (N=37)		Spontaneous speech (N=37)	
	Mean	SD	Mean	SD	Mean	SD
F0mean	115.1	15.2	112.1	12.2	109.1	13.6
F1	422.4	44.7	496.6	39.1	535.9	53.0
F2	1751.5	96.5	1382.8	62.2	1643.1	63.6
F3	2568.3	78.3	2471.2	84.4	2576.7	86.2
F4	3461.9	151.7	3439.8	150.5	3548.5	132.9
ΔF	1017.9	35.9	978.1	32.3	1028.3	30.4

Perceived Masculinity

Each of the 20 female raters was sat in a sound-controlled room in front of a computer screen and wore *Dynamode dh-660mv* headsets. Raters were able to adjust the sound volume to a comfortable level prior to the rating task. Voice stimuli for the 37 male speakers were presented using a custom script written in *PRAAT v.5.2.17* in three separate blocks according to stimuli type (isolated words, sentence and spontaneous speech), and with stimuli order randomised within each block. After listening to each stimulus, listeners were asked to rate the speaker's masculinity ("how masculine does the speaker sound?") on a seven-point Likert scale from 1 (not at all masculine) to 7 (very masculine), by clicking on one of the equally sized buttons labeled from 1 (left endpoint) to 7 (right endpoint). Each rater thus made 111 judgments (37 speakers x 3 blocks), with scheduled rest-breaks every 13 stimuli. Mean ratings across listeners are reported in Table 3.3.3.

Table 3.3.3

Ranges, means and standard deviations for perceived masculinity ratings of male speakers from voice stimuli (isolated words, sentence, and spontaneous speech)

Voice stimuli	N	Range	Mean perceived masculinity rating	SD
Isolated words	37	1–7	4.94	1.47
Sentence	37	1–7	4.85	1.49
Spontaneous speech	37	1–7	5.05	1.59

Modelling Analysis

In order to test multiple pathways from biological (testosterone, height) and acoustic (F0, ΔF) characteristics of speakers to masculinity ratings of their voices, we run a two-level (speaker nested within listeners) path analysis for each of the three perceptual tasks on the fully saturated model (shown in Figure 3.3.1). The speakers' sample size was a little over the recommended minimum of 5 cases per parameter (Lei & Wu, 2007) necessary to perform this analysis. Standardised path coefficients (ρ) and their significance levels, as well as indirect and total effects, were calculated with *Mplus v.7.11* (Muthén & Muthén , 2013) using the ML (maximum likelihood) estimator. The strength of the associations was interpreted following Campbell & Swinscow (1996): values of ρ .00–.19 are regarded as very weak, .20–.39 as weak, .40–.59 as moderate, .60–.79 as strong and .80–1.00 as very strong. R^2 values showed that the model accounted for a relatively small percentage of the variance in perceived masculinity in each task (word: 17.2%, sentence: 28.4%, speech: 26.1%). Standardised estimates are reported in Figure 3.3.1.

Inter-rater reliability was calculated using Cronbach's alpha (*α)* for each of the three voice tasks (isolated words: α = .937, sentence: α = .928, spontaneous speech: α = .933). Agreement between raters with respect to the actual values they assigned individuals was assessed by the Intra-class Correlations (ICC) from the multilevel analysis (speakers nested within listeners) (isolated words: ICC = .27; sentence: ICC = .20; spontaneous speech: ICC = .22). Since the degree of reliability of ratings among participants was high (α > 0.8 in all cases) and agreement was fair (ICC = .21–.40, Jewell, 2011), we consider that in general female listeners agreed on their masculinity ratings.

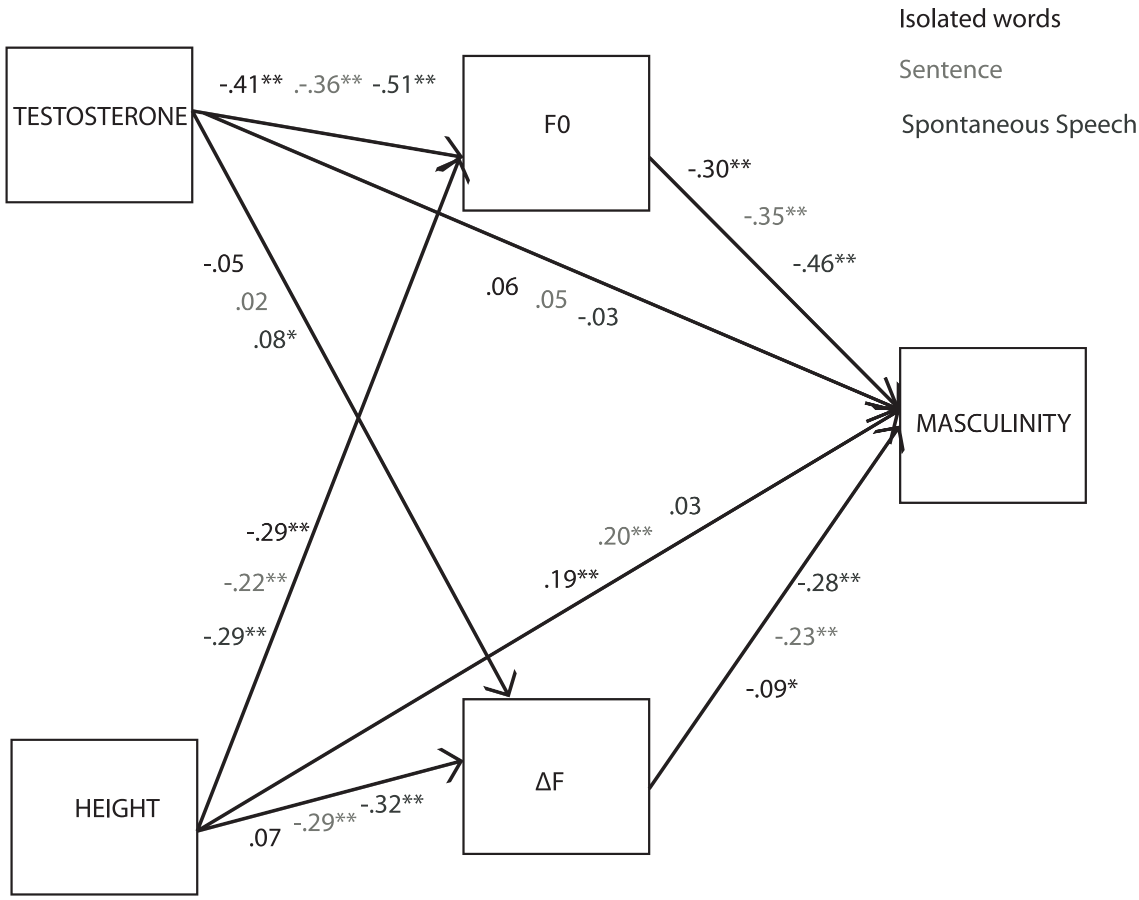

Figure 3.3.1. Path diagram showing path coefficients for each of the three tasks (word in black, sentence in light-grey and speech in grey). Residuals for meanF0 and ΔF are allowed to vary. Significant coefficients $p < .05$, $p < .001$ are reported with asterisks *, ** respectively.

Results

Biological and Acoustic Characteristics

Testosterone and height were significantly, though weakly, correlated ($\rho = -.25$, $p < .001$). Men with higher salivary testosterone levels were found to have lower F0 (lower pitch) in all three tasks (ps < .001), and the correlation between the two variables was moderate ($\rho = -.36$ to $-.51$). Men with higher testosterone levels had significantly higher ΔF in the speech task only, though the correlation was very weak ($\rho = .08$, $p = .028$). A weak and yet significant correlation was found between body height and ΔF in the sentence and speech tasks ($\rho = -.29$ to $-.32$, ps < .001), with taller men having lower ΔF (deeper timbre). Taller men also had significantly lower F0, though the correlation between the two variables was weak in all tasks (ρ = -.22 to -.29, *ps* < .001).

Acoustic Characteristics of the Speakers and Listeners' Judgments

F0 and ΔF were significantly, but very weakly, correlated in all three tasks (ρ = .17 to .18, *p*s < .001). Men with lower F0 were perceived as more masculine in all three tasks (ρ = -.30 to -.46, *p*s < .001). Men with narrower ΔF were also perceived as more masculine in all three tasks, though path coefficients revealed that ΔF had a weaker correlation with perceived masculinity than F0 across tasks (ρ = -.09 to -.28, *ps* < .05).

Biological Characteristics and Listeners' Judgments

Taller, higher-testosterone men were perceived as more masculine in all tasks. The total effect sizes for the paths from height to perceived masculinity were bigger (ρ = .25–.34, *p* < .001) than those found for the paths from testosterone to perceived masculinity (ρ = .17–.19, *ps* < .001), indicating that height was more strongly correlated with perceived masculinity than testosterone. Inspection of the effect sizes for the indirect and direct paths from testosterone to perceived masculinity revealed that the relationship between the two variables was almost entirely mediated by F0 (ρ = .12–.23, *ps* < .001), though a very small, yet significant indirect path in the opposite direction was found in the speech task via ΔF (ρ = -.002, *p* < .05), while the direct path from testosterone to perceived masculinity was not significant (*ps* > .05). With the exception of speech (*p* > .05), the direct paths between height and perceived masculinity were significant and their effect sizes greater (ρ = .19–.20, *ps* < .001) than those of the indirect paths between the two variables (ρ =.08–.14, *ps* < .05), revealing that the relationship between height and perceived masculinity was entirely mediated by F0 and ΔF in the speech task only.

Discussion

These results reveal clear associations between fitness-related characteristics, sexually dimorphic acoustic traits, and perceived masculinity: individuals who are taller and have higher testosterone levels tend to speak with lower fundamental frequency and lower formant frequency spacing, and tend to be rated as more masculine from their voice by female listeners.

Biological Characteristics and Voice Cues

In line with our hypotheses and previous research (Dabbs & Mallinger, 1999; Evans et al., 2008; Puts et al. 2012), speakers' salivary testosterone was negatively correlated with their voice F0. Also in line with recent research (Evans et al., 2008; Puts et al., 2012; though see Bruckert et al., 2006), salivary testosterone level was not a significant predictor of ΔF, with the exception of a very weak ($\rho = .08$), yet significant ($p = 0.028$) association between the two measures in spontaneous speech (in the unexpected direction: higher testosterone men spoke with marginally higher ΔF).

While our observations support stronger associations of testosterone with F0 than with ΔF, the examination of the relationship between testosterone and vocal parameters in adulthood remains incomplete. Longitudinal studies would help clarify whether individual differences in acoustic features linked to testosterone reflect variance in total testosterone exposure during (pubertal) development or a more gradual, continuing exposure spanning across adulthood. So far, evidence from androgen treatment of individuals lacking the masculinisation of the larynx (e.g. female-to-male transsexuals and adult males with hypogonadism) has shown that vocal folds are still sensitive to testosterone in adulthood, with testosterone injections permanently thickening the folds and thus lowering voice F0 (Akcam et al., 2004; Baker, 1999; Talaat et al., 1987; Van Borsel et al., 2000). However, the potential effects of testosterone exposure in adult males without androgen deficiencies on the physiology (vocal fold mass and length, vocal tract length and extensibility) and on the behavioural control of the vocal apparatus (Pisanski et al., 2012) remain largely unknown.

We also reported a negative, weak, and yet significant correlation between height and F0 across all tasks (*ps* < .001), with taller men speaking with lower F0. While F0 accurately cues for body size between sex and age classes (adult men have lower F0 than women and children (Titze, 1994), its role as a predictor of body size within-sex remains equivocal. The weak relationship between F0 and body height is consistent with the absence of skeletal structures constraining the dimensions of the larynx, which results in vocal fold length being largely unrelated to overall body size

(Fitch, 2000). Indeed, in line with the present study, most acoustic studies have reported a weak correlation (Graddol & Swann, 1983; Puts et al., 2012) or no correlation between F0 and height (Evans et al., 2006; González, 2004; Künzel, 1989; Lass, 1978; Rendall, 2005; Sawashima et al., 1983; Sell et al., 2010; but also see Graddol & Swann, 1983; Puts et al., 2012). Despite being a poor cue to speaker size, psychoacoustic studies have consistently reported the perceptual salience of F0 in size ratings (Rendall, et al., 2007; Van Dommelen 1993; Feinberg et al., 2005; Fitch, 1994; Pisanski & Rendall, 2011; Smith & Patterson, 2005), leading several authors to suggest that listeners may overgeneralise between-sex and age differences (Rendall, et al., 2007), or broader sound-size associations in the natural world (e.g. large objects producing bass sounds (Grassi, 2005; Rendall, et al., 2007)). The present methodology could be usefully replicated with listeners' ratings of body size to further investigate the three-way relationship between voice cues, actual and perceived body size.

Also in line with our hypotheses, we found that taller men spoke with narrower formant spacing, though the association between the two measures was weaker than the one reported between testosterone and F0. Unlike the larynx, the length of the vocal tract is relatively more constrained by the skeletal anatomy that surrounds it (neck and skull), which is in turn affected by overall body size (Rendall, et al., 2007). Therefore ΔF, the inverse acoustic correlate to vocal tract length, may also provide a reliable cue to body size and in particular height (Fitch & Giedd, 1999). Most acoustic studies have indeed found moderate correlations between ΔF and men's height (Bruckert, et al., 2006; Evans et al., 2006; Greisbach, 2007; Puts et al., 2012; Rendall et al., 2005; Rendall, et al., 2007; Sell et al., 2010), though others have failed to find correlations between the two measures (Collins, 2000; Van Dommelen & Moxness, 1995).

Voice cues and Listeners' Ratings

We found that men speaking with lower frequency values attracted higher masculinity ratings. These results are consistent with psychoacoustic studies showing that male voices characterised by lower F0, lower ΔF, or both (Feinberg et al., 2008; Pisanski & Rendall, 2011; Pisanski et al., 2012) receive higher masculinity ratings by women (and male) listeners than those with the same parameters raised, lending further support to the hypothesis that women attend to sexually-dimorphic, androgen-dependent voice characteristics when assessing value of potential mates. Furthermore, in the present study F0 was a more salient cue for perceived masculinity than ΔF. One explanation is that women listeners weighed F0 cues more heavily than ΔF cues when rating the perceived masculinity of speakers because of a stronger link between F0 and underlying fitness-related features compared to ΔF. Alternatively, listeners may simply take advantage of the greater sex-dimorphism in F0 compared to ΔF when assessing speakers' gender-related traits. Indeed, while, in natural voices, F0 appears to be a more salient cue to speakers' sex and masculinity than ΔF (Collins, 2000; Hillenbrand & Clark, 2009), this salience is in fact reversed when the magnitude of variation is controlled by making the two cues equally perceptually discriminable (Pisanski & Rendall, 2011).

To what Extent do F0 and ΔF mediate the Relationship between Size, Androgens and Perceived Masculinity?

Higher testosterone levels were associated with higher masculinity ratings and this relationship was fully mediated by F0 (except for a marginal mediatory role of ΔF in the speech task), with higher testosterone men having lower F0 and in turn being rated as more masculine. We also observed a mediatory role of both F0 and ΔF in the relationship between perceived masculinity and height, with taller men having lower F0 and ΔF and, in turn, being attributed higher masculinity ratings. The mediatory effects of these two acoustic cues were similar in magnitude across tasks. Additionally, the significant relationships between height and perceived masculinity were still present when the mediatory effects of the acoustic cues were partialed out (except for speech), suggesting that height affects perceived masculinity via additional voice cues. The availability of extra cues to height may account for the marginally stronger relationship between height and perceived masculinity than between testosterone and perceived masculinity. Studies with read-aloud and spontaneously uttered speech have also highlighted the role of cues other than F0 and ΔF in the expression and perception of voice masculinity, such as prosody (Cartei & Reby, 2012; Cartei et al., 2012; John-Lewis, 1986).

While confirming the mediatory role of F0 and ΔF between biological and perceptual measures of masculinity across all speech tasks (isolated words, sentence and spontaneous speech), the present study also shows that the strength of the relationship tends to increase from the shortest and least naturalistic stimuli (isolated monosyllabic words) to the longest and more naturalistic and ecologically valid stimuli (spontaneous speech). This suggests that studies investigating the function of nonverbal voice cues in human interactions should use realistic stimuli and interpret results from isolated vowels or words with caution.

Moreover, it is important to note that our study used speech material with relatively neutral content. Vukovic and colleagues (2010) have found that positively valenced men's speech (e.g. 'I really like you' as opposed to 'I don't really like you'), increases women's preferences for masculinised voices, suggesting in turn that semantic content (at least when expressing mating interest) may affect the links between biological, acoustic and perceptual dimensions. Further studies should investigate whether the correlations we report may be accentuated by the use of speech material with a content highlighting the relevance of masculinity (e.g. dating related).

Conclusions

This study expands on previous investigations of masculinity expression in the human voice, by explicitly exploring the relationships among biological (body height and testosterone), acoustic (F0 and ΔF) and perceptual dimensions (women listeners' ratings) of males' masculinity. While the overall results of this study confirm links among all three dimensions, the observed variation in the mediatory effects of F0 and ΔF between the biological and perceptual dimensions warrants future research. For example, future investigations should take into account listeners' individual differences, such as women's fertility (Feinberg et al., 2006), body size (Feinberg et al., 2005) and self-rated attractiveness (Vukovic et al., 2008), which have been shown to differentially effect women's preferences of males' voices.

Moreover, while salivary testosterone is commonly used as a biological marker of masculinity because of its relative temporal stability (Dabbs, 1990a; Sellers et al., 2007), it has also been shown to vary daily and seasonally (Dabbs, 1990b), and in response to different social contexts (e.g. increasing after 'winning' (Booth et al., 1989)). Thus, replication and extension of the current findings, preferably with repeated testosterone assays to account for testosterone variations, and the inclusion of additional correlates of masculinity (e.g. facial width-to-height ratio: Lefevre et al., 2013; reproductive success: Apicella et al., 2009), would be desirable to shed light on the extent to which acoustic features cue for fitness-related traits.

Finally, adults have been found to spontaneously modify sex-dimorphic acoustic cues (F0 and ΔF) in order to vary the expression of gender and related attributes in line with different roles and social (e.g. gender expression, dominance, sexual orientation) contexts (Cartei et al., 2012; Cartei & Reby, 2012; Graddol & Swann, 1983; Puts et al., 2007), and this variation has a strong effect on the way listeners perceive the personality of speakers (Owen et al., 2010; Puts et al., 2006; Van Bezooijen, 1995). Because variation in vocal masculinity is likely to have both biological and social sources, future studies should also include social measures of masculinity (e.g. speakers' self-ratings of masculinity) in order to further explore how vocal masculinity relates to speakers' characteristics (both biological and social) and how these are perceived by listeners.

References

Akcam, T., Bolu, E., Merati, A. L., Durmus, C., Gerek, M., & Ozkaptan, Y. (2004). Voice changes after androgen therapy for hypogonadotrophic hypogonadism. *The Laryngoscope*, *114*(9), 1587–1591.

Andersson, M. B. (1994). *Sexual Selection*. Princeton University Press.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, *3*(6), 682–684.

Baker, J. (1999). A report on alterations to the speaking and singing voices of four women following hormonal therapy with virilizing agents. *Journal of Voice*, *13*(4), 496–507.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2.17)[Computer program]. Retrieved January 1, 2012.

Booth, A., Shelley, G., Mazur, A., Tharp, G., & Kittok, R. (1989). Testosterone, and winning and losing in human competition. *Hormones and Behavior*, *23*(4), 556–571.

Bruckert, L., Liénard, J.-S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 83–89.

Campbell, M. J. & Swinscow, T. D. V. (1996). *Statistics at Square One*: *Ninth Edition*. London: BMJ Publ. Group,

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous Voice Gender Imitation Abilities in Adult Speakers. *PLoS ONE*, *7*(2), e31353.

Cartei, V., & Reby, D. (2012). Acting Gay: Male Actors Shift the Frequency Components of Their Voices Towards Female Values When Playing Homosexual Characters. *Journal of Nonverbal Behavior*, *36*(1), 79–93.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*(6), 773–780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*(5), 997–1004.

Dabbs, J. M., Jr. (1990a). Salivary testosterone measurements: reliability across hours, days, and weeks. *Physiology & behavior*, *48*(1), 83–86.

Dabbs, J. M., Jr. (1990b). Age and seasonal variation in serum testosterone concentration among men. *Chronobiology international*, *7*(3), 245–249.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and individual differences*, *27*(4), 801–804.

Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biological psychology*, *72*(2), 160–163.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4), 783–788.

Fairbanks, G. "The rainbow passage." *Voice and Articulation Drill Book* (1960): 124-139.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior*, *29*(4), 233–241.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G. & Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, *49*(2), 215–222.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568.

Fink, B., Neave, N., & Manning, J. T. (2003). Second to fourth digit ratio, body mass index, waist-to-hip ratio, and waist-to-chest ratio: their relationships in heterosexual men and women. *Annals of human biology*, *30*(6), 728–738.

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology*, *19*(1), 82–87.

Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language*. Brown University.

Fitch. W.T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences*, *4*(7), 258–267.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*, 1511.

Gallup, A. C., White, D. D., & Gallup Jr, G. G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior*, *28*(6), 423–429.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology*, *6*(4), 219–225.

Graddol, D., & Swann, J. (1983). Speaking Fundamental Frequency: Some Physical and Social Correlates. *Language and Speech*, *26*(4), 351–366.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & psychophysics*, *67*(2), 274–284.

Greisbach, R. (2007). Estimation of speaker height from formant frequencies. *International Journal of Speech Language and the Law*, *6*(2), 265–277.

Harries, M., Hawkins, S., Hacking, J., & Hughes, I. (1998). Changes in the male voice at puberty: vocal fold length and its relationship to the fundamental frequency of the voice. *The Journal of Laryngology & Otology*, *112*(05), 451–454.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166. doi:10.3758/APP.71.5.1150

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2011). Voice correlates of mating success in men: Examining “contests” versus “mate choice” modes of sexual selection. *Archives of Sexual Behavior*, *40*(3), 551–557.

Hönekopp, J., Rudolph, U., Beier, L., Liebert, A., & Müller, C. (2007). Physical attractiveness of face and body as indicators of physical fitness in men. *Evolution and Human Behavior*, *28*(2), 106–111.

Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. *Intonation in discourse*, 199-220.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57–62.

Lass, N. J. (1978). Correlational study of speakers’ heights, weights, body surface areas, and speaking fundamental frequencies. *The Journal of the Acoustical Society of America*, *63*(4), 1218–1220.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, *34*(4), 273–279. doi:10.1016/j.evolhumbehav.2013.03.005

Lei, P. W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33-43.

Little, A. C., Jones, B. C., & Burriss, R. P. (2007). Preferences for masculinity in male bodies change across the menstrual cycle. *Hormones and Behavior*, *51*(5), 633–639.

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1496), 1095–1100.

Llisterri, J. (1992). Speaking styles in speech research. In *Workshop on Integrating Speech and Natural Language*.

Künzel, H. J. (1989). How Well Does Average Fundamental Frequency Correlate with Speaker Height and Weight? *Phonetica*, *46*(1-3), 117–125.

Muthén, L. K.., & Muthén, B. O. (2013). MPLUS (Version 7.11). [Computer Software].

Los Angeles, CA.

Owen, K., & Hancock, A. B. (2010). The Role of Self- and Listener Perceptions of Femininity in Voice Therapy. *International Journal of Transgenderism*, *12*(4), 272–284.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. Evolution and Human Behavior, *25*, 229–241.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, *33*(5), 509–519.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*, 2201.

Prokop, P., & Fedor, P. (2013). Associations between body morphology, mating success and mate preferences among Slovak males and females. *Anthropologischer Anzeiger*, *70*(2), 121–135.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*(5), 388–397.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 601–609.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*, 283-296.

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature communications, 3*, 694.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*, *117*(2), 944–955.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *Journal of Experimental Psychology-Human Perception and Performance*, *33*(5), 1208–1219.

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings. Biological sciences / The Royal Society*, *270 Suppl 1*, S93–95.

Sawashima, M., Hirose, H., & Honda, K. 1983. Stereoendoscopic measurement of the laryngeal structure. In: Bless, H.H. (Ed.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. College-Hill, New York

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., Krauss, A. & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3509-3518.

Sellers, J. G., Mehl, M. R., & Josephs, R. A. (2007). Hormones and personality: Testosterone as a marker of individual differences. *Journal of Research in Personality*, *41*(1), 126–138. doi:10.1016/j.jrp.2006.02.004

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*, 3177.

Talaat, M., Angelo, A., Talaat, A. M., Elwany, S., Kelada, I., & Thabet, H. (1987). Histologic and histochemical study of effects of anabolic steroids on the female larynx. *The Annals of otology, rhinology & laryngology*, *96*(4), 468–471.

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*(2), 131–144.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and speech*, *38 (Pt 3)*, 253–265.

Van Borsel, J., De Cuypere, G., Rubens, R., & Destaerke, B. (2000). Voice problems in female-to-male transsexuals. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, *35*(3), 427–442.

Van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and speech*, *38 (Pt 3)*, 267–287.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., & Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: an imaging study. *The Journal of the Acoustical Society of America*, *125*(3), 1666–1678.

Vukovic, J., Feinberg, D. R., Jones, B. C., DeBruine, L. M., Welling, L. L. M., Little, A. C., & Smith, F. G. (2008). Self-rated attractiveness predicts individual differences in women's preferences for masculine men's voices. *Personality and Individual Differences*, *45*(6), 451–456.

Vukovic, J., Jones, B. C., DeBruine, L., Feinberg, D. R., Smith, F. G., Little, A. C., Welling, L. L. M., & Main, J. (2010). Women's own voice pitch predicts their preferences for masculinity in men's voices. *Behavioral Ecology*, *21*(4), 767–772.

Welling, L. L. M., Jones, B. C., DeBruine, L. M., Conway, C. A., Law Smith, M. J., Little, A. C., Feinberg, D. R., Sharp, M. A., & Al-Dujaili, E. A. S. (2007). Raised salivary testosterone in women is associated with increased attraction to masculine faces. *Hormones and Behavior*, *52*(2), 156–161.

Zebrowitz, L. A., & Rhodes, G. (2002). Nature let a hundred flowers bloom: The multiple ways and wherefores of attractiveness. In *Facial attractiveness: Evolutionary, cognitive and social perspectives* (pp. 261–293). Westport, CT.: Ablex.

Chapter 4:

Sexually Dimorphic Cues Are Behaviourally Altered to Vary Voice Gender Expression

Summary

The previous chapter confirmed that sexually dimorphic cues in the voice (ΔF in pre-pubertal children and adults, and F0 in adults) signal gender and related attributes (e.g. masculinity), and provided some evidence that acoustic variation of these cues must be possible within the physical constraints of speakers' vocal apparatus and is linked to gender-typed vocal behaviours.

The main aim of this chapter is to further investigate the latter hypothesis by exploring whether speakers exploit the acoustic variation in sexually dimorphic voice cues to underplay or accentuate gender (maleness, femaleness) and related attributes (masculinity, femininity).

More specifically, the following questions will be explored:

Question 3. Can individuals control fundamental and formant frequencies in order to vary the expression of gender, masculinity and femininity of their voice, and does the acoustic co-variation of these parameters occur along the existent sex dimorphism?

Question 4. Are speakers aware of what voice and articulatory gestures they use to vary the gender expression in their voice?

Question 5. What is the perceptual relevance of these gestures?

Study 4 investigates Question 3 in relation to pre-pubertal child speakers, by asking six- to nine- year olds to sound "like a boy" or "like a girl" as much as possible (while reading words out loud), and testing whether they would decrease or increase ΔF (in line with the observed acoustic dimorphism). F0 was also measured, though variation in this parameter was not expected (as not sexually dimorphic in pre-pubertal voices). Children were also asked about what they did to spontaneously vary the expression of their voice gender (Question 4 – see addition to Study 4 (4.1)).

Summary of findings:

- Pre-pubertal children adjusted their ΔF when asked to alter the gender of their voice, exaggerating behavioural differences in ΔF that exist in their age group

- When imitating the opposite gender, pre-pubertal children also mimicked biological differences in F0 that exist in adults (despite no differences at their age)
- When asked to describe how they achieved the target voice, children spontaneously focused on the perceptual outcome (e.g. making their voice ‘higher’, or ‘lower’) rather than the production gestures used. Moreover, girls were aware of lowering their voice when sounding like a boy, while both boys and girls showed some awareness of raising their voice when sounding like a girl
- When given a choice of possible gestures, boys and girls did not report glottal (changes in pitch) or vocal tract (via lip spreading or laryngeal vertical shifts) adjustments

Study 5 investigates Question 3 in relation to adult speakers, by asking individuals to sound “as masculine” or “as feminine” as possible (while reading words, sentence and a passage out loud) and testing whether they would decrease or increase their F0 and ΔF (in line with the observed post-pubertal acoustic dimorphism in both parameters). Question 4 was also explored by asking speakers to describe what they did to spontaneously vary the expression of their voice gender, and by quantitative measurements of lip movements.

Summary of findings:

- Adult male and females adjusted their F0 and ΔF along the existing, biologically determined dimorphism when asked to alter the masculinity and femininity of their voice
- Women displayed greater lip spreading and opening than men suggesting that women spoke with a “smile”. However, both men and women moved their lips in a similar way across all three conditions
- When asked to describe how they achieved the target voice, speakers spontaneously focused on the perceptual outcome (e.g. making their voice ‘higher’, or ‘lower’) rather than the production gestures used

- When given a choice of possible gestures, speakers showed greater awareness of glottal rather than vocal tract adjustments. Moreover, men showed greater awareness of larynx lowering than women when masculinising their voices

Study 6 investigates whether behavioural adjustments of F0 and ΔF are salient and relevant when assessing speakers' perceived gender (Question 5), by asking adult listeners to make gendered attributions from normal, masculinised and feminised voices of pre-pubertal children (from Study 4) and adult speakers (from Study 5).

Summary of findings:

- Masculinised (lower F0 and ΔF) adult voices were described as more masculine, and feminised adult voices (higher F0 and ΔF) as less masculine, than adults' normal speaking voices
- Boys' normal voices were rated as significantly more masculine than girls', while their masculinised and feminised voices received similar ratings to girls'
- Men's voices were rated as significantly more masculine than women across conditions

Study 4:

Control of Voice Gender in Pre-pubertal Children

Note. Study published as: Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2013). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology, 32*(1).

Abstract

Adult listeners are capable of identifying the gender of speakers as young as four years old from their voice. In the absence of a clear anatomical dimorphism in the dimensions of pre-pubertal boys' and girls' vocal apparatus, the observed gender differences may reflect children's regulation of their vocal behaviour. A detailed acoustic analysis was conducted of the utterances of 34 six- to nine- year old children, in their normal voices and also when asked explicitly to speak like a boy or a girl. Results showed statistically significant shifts in fundamental and formant frequency values towards those expected from the sex-dimorphism in adult voices. Directions for future research on the role of vocal behaviours in pre-pubertal children's expression of gender are considered.

Introduction

Introducing a recent special issue on gender and relationships, Leman and Tenenbaum (2011, p. 153) draw attention to "the ways in which children practise future gender roles in everyday interactions with their peers and parents." Indeed, children are known to exhibit gender-typed patterns of behaviour from a young age. Boys and girls prefer gender-normative toys (Martin, Eisenbud & Rose, 1995) and play styles (Munroe & Romney, 2006; Hay et al., 2011), and are more likely to choose same-sex peers as playmates (Golombok et al., 2008; Zosuls et al., 2011). We also know that young children are capable of regulating their behaviour in gender-typed ways – what we might call 'self-presentation of gender' – under given social circumstances, such as the presence of a same-sex peer group (Banerjee & Lintern, 2001). With regard to verbal behaviour, much attention has been paid to the content, style, language use, and social dynamics of boys' and girls' conversations (e.g. Leaper & Smith, 2004; Leman, Ahmed, & Ozarow, 2005). Yet, surprisingly, one of the most obvious aspects of gender difference in verbal interactions – the voice itself – has been largely ignored.

Adults can identify the gender of speakers as young as four years of age by listening to their voice only (Perry et al., 2001). In post-pubertal speakers, sex differences in the dimensions of the vocal apparatus give males a lower fundamental frequency (pitch) and lower vocal tract resonances (or formants). Before puberty, boys also speak with lower vocal tract resonances than girls (but with the same pitch: Perry et al., 2001). However, these acoustic differences are not supported by a corresponding anatomical sex-dimorphism, suggesting that they have a strong behavioural dimension: children seem to adjust the length of their vocal tract to produce formant frequencies characteristic of their gender. See Appendix A for details on sex dimorphism in the human voice.

The hypothesis that children control this aspect of their vocal behaviour is plausible in light of empirical research showing that children from a young age make use of the voice, along with other cues such as faces, in discriminating males and females (see Ruble, Martin, & Berenbaum, 2006). The expression of voice gender is therefore a very promising and objectively quantifiable indicator of gender development in children. So far, though, children's ability to control the gender-related characteristics of their voices has never been directly investigated.

We report here on the ability of six- to nine- year old (pre-pubertal) children to shift the frequency components of their voices when they are prompted to alter their perceived gender. Using a paradigm that has previously been successful in revealing adults' ability to control gender-typed acoustic parameters (Cartei et al., 2012), we asked children to sound "like a boy" or "like a girl" as much as possible and evaluated their capacity to control fundamental frequency and formant frequencies (decreasing their spacing to sound more like a boy, and increasing it to sound more like a girl).

Method

Participants

Voice recordings were obtained from 34 children (15 boys and 19 girls), aged six to nine, M = 7.04, SD = 1.11 (see Table 4.4.1 for detailed age and sex distribution of participants). The children had no history of hearing or speech impediments and were all native speakers of British English. Height and weight were measured for each child, and no sex differences were found, ps > .10.

Table 4.4.1

Distribution of male and female speakers

Age (years)	Sample size	Males	Females
6	14	4	10
7	8	5	3
8	6	4	2
8.5	3	1	2
9	3	1	2

Procedure

Recordings were made of the children in one-to-one interactions with the experimenter, in a quiet room at the child's school or at a university laboratory. All audio recordings were made using a *Tascam DR07mkII* handheld recorder connected to a *Shure SM94* microphone. Each participant was shown nine cards with a written and pictorial representation of the target words (e.g. the image of a bed and underneath the word "bed"), and asked to say the words on the cards, first in their normal speaking voice (the instruction was: "please read these words out loud"), then trying to sound as much as possible "like a boy" or "like a girl", in alternate order (the instruction was: "now please read these words out loud trying to sound like a boy (or a girl) as much as possible"). The order in which the cards were presented was randomized across participants to avoid serial order effects.

Acoustic Analyses

The speech material consisted of nine non-diphthong vowels of British English embedded in CVC words (/ae/ "hat", /eh/ "bed", /er/ "bird", /iy/ "feet", /ih/ ”pig", /ah/ "duck", /aa/ "box", /uh/ "book", /uy/ "boot"). All acoustic analyses were conducted on the steady portion of each vowel, with PRAAT *v.5.2.17* (Boersma & Weenink, 2011) using a custom written script for batch processing (available from the authors on request).

The script calculated the mean fundamental frequency (F0), the perceptual correlate of voice pitch, with lower F0 resulting in lower-pitched voices. Additionally, the script estimated the centre frequencies of the first four formants (F1–F4) of each vowel. The difference between any two adjacent formant frequencies, also defined as formant spacing, was then calculated ($\Delta F = F_{i+1} - F_i$) and used for analysis as this gives a more accurate estimate of global vocal tract adjustments than individual formant values. Longer vocal tracts produce lower formant spacing, giving voices a more baritone quality (see Appendix B for details of acoustic analyses and Appendix C for descriptive statistics for a wider range of acoustic parameters).

Results

Table 4.4.2 summarises the mean values and standard deviations for fundamental frequency and formant spacing in the three conditions.

Table 4.4.2

Mean (SD) in Hz for fundamental frequency (F0) and format spacing (ΔF) of boys and girls in the masculinised, natural and feminised conditions

Sex of Speaker	Acoustic Parameter	Masculinised	Natural	Feminised
Boys	F0	243.5 (32)	249.6 (29)	307.2 (62)
	ΔF	1284 (69)	1313 (68)	1355 (80)
Girls	F0	234.6 (30)	249.1 (26)	270.2 (50)
	ΔF	1301 (67)	1355 (46)	1389 (53)

Age and Sex Differences in the Natural Voice

We first performed a series of ANCOVAs in order to test the effects of sex and age (continuous covariate) on the acoustic parameters F0 and ΔF of children's natural voices. There was a significant effect of age on mean F0, with F0 decreasing as children get older, $F(1, 34) = 4.88$, $p = .035$. No significant main effect of sex was found, $F(1, 34) = .07$, $p > .10$. There was a main effect of sex on children's natural ΔF, with boys speaking with a 43Hz lower ΔF than girls, $F(1, 34) = 4.23$, $p = .048$. There was a non-significant tendency of ΔF to decrease with age, $F(1, 34) = 3.95$, $p = .056$.

Ability to Control Voice Gender

We assessed the ability of boys and girls to shift different acoustic parameters by testing the main effect of condition (three-level within-subject factor: natural, masculinised, feminised) on the acoustic parameters with repeated measures ANOVA within each sex. We also investigated whether any of the shifts between natural voices and the two conditions were significantly associated with age by calculating the difference between the natural and masculinised or feminised conditions and regressing these difference variables on age.

The ANOVAs on F0 showed that the main effect of condition was significant in boys, $F(1.18, 16.48) = 14.09$, $p = .001$, and in girls, $F(1.22, 21.94) = 6.93$, $p = .011$. Within-sex contrasts revealed that, when asked to sound as much like a boy as possible, boys did not significantly lower F0 compared to the natural condition, $F(1, 14) = 1.04$, $p > .10$. In contrast, when feminising their voices, they significantly raised their F0 by 23.2% (3.59 ST) from 249.6Hz to 307.2Hz. $F(1, 14) = 16.18$, $p = .001$. Simple regression revealed that the magnitude of this upward shift increased with age ($R^2 = .34$, $F(1, 14) = 6.68$, $\beta = .58$, $p = .023$). Correspondingly, girls significantly lowered their F0 by 5.8% (1.04 ST) from 249.1Hz to 234.6Hz, $F(1, 18) = 10.11$, $p = .005$, when masculinising their voices, but did not significantly raise F0 when feminising them, $F(1, 18) = 3.09$, $p = .096$. Age was not significantly related to girls' F0 difference scores, *βs* = -.01 and .19, *ps* > .100.

The corresponding ANOVAs for ΔF showed that condition had a significant effect in both boys, $F(1.18, 16.54) = 16.35$, $p = .001$, and girls $F(2, 36) = 24.19$, $p < .001$. Within-sex contrasts revealed that both sexes significantly lowered ΔF (by 2.2% in boys, $F(1, 18) = 31.63$, $p < .001$, and by 3.9% in girls, $F(1, 18) = 20.21$, $p < .001$) to sound more masculine and significantly raised it to sound more feminine (by 3.2% in boys, $F(1, 14) = 8.20$, $p = .013$, and 2.5% in girls, $F(1, 18) = 10.48$, $p = .005$). No significant associations were found between ΔF difference scores and age, *β*s = -.04 to .27, *p*s >.100.

Discussion

Our analyses confirmed that boys displayed narrower formant frequency spacing than girls in their natural voice (Perry et al., 2001), and revealed that speakers of both sexes shifted this parameter along the existing sex dimorphism when asked to alter their voice gender. They also revealed that, despite the confirmed absence of sex differences in the fundamental frequency of pre-pubertal children's natural voices, both boys and girls adjusted this parameter when imitating the opposite sex in line with the sex differences present in adults.

Given the absence of sex differences in overall anatomical vocal tract length before puberty (Fitch & Giedd, 1999; Vorperian et al, 2011), sex differences in formant spacing suggest that children behaviourally adjust their vocal tract length via lip protrusion (or spreading) and/or larynx lowering (or raising) to advertise their gender in their natural voice. The fact that children further control this parameter when altering the gender of their voice provides tentative support for this hypothesis: both sexes lowered their formant spacing to masculinise their voice and raised them to feminise it, as previously observed in adults (Cartei et al., 2012). While the vocal tract adjustments observed here are only temporary, and in response to an explicit request, they nevertheless provide the first evidence that children have the ability to manipulate these acoustic properties in order to achieve gender-typed voices. The specific nature of the articulatory gestures involved could be studied more directly using cine-MRI.

The role of F0 in the expression of voice gender appears to be more nuanced. In the natural voice condition, F0 was not significantly different between boys and girls, consistent with most acoustic data (Lee et al., 1999; Sachs et al., 1973) and with the absence of sex dimorphism in the development of vocal fold and laryngeal morphology reported by previous anatomical studies (Kahane, 1978; Titze, 1994). This suggests that F0 may not play a role in advertising gender in pre-pubertal children's voices when they are in a neutral context. However, children lowered their mean F0 when asked to masculinise their voices, whereas they raised it when feminising their voices. The shifts of F0 were significant when children were asked to sound like the opposite gender, in line with what was previously reported in adults (Cartei et al., 2012). Evidently, children have (at least implicitly) some knowledge of adult sex differences in F0 and may use it to vary the gender of their voice. Moreover, and notwithstanding our relatively small and gender unbalanced sample, there was evidence that boys' manipulation of F0 to feminise their voices increased with age. Interestingly, children did not significantly shift F0 to exaggerate their own gender, in contrast with observations in adults (Cartei et al., 2012). Further studies with a larger, more balanced sample, across a wider age range, are warranted to confirm these results and further investigate the use of F0 to express gender in line with age and gender differences. In addition, our study was limited by its reliance on assessing single-word vocal production within a restricted laboratory context; future research can fruitfully target children's natural speech in different settings.

Self-Presentation of Gender through the Voice?

The "size-code" hypothesis (Ohala, 1984), which predicts that callers make a conventionalised use of primarily size-related acoustic variation to communicate motivational information, has received support from both non-human (Reby et al., 2005) and human (Puts et al., 2006) studies showing that males lower their frequency components to sound more dominant. We propose that, because in humans F0 and ΔF are primarily indexes of sex rather than size, speakers primarily use a "gender code", whereby they control these cues to vary the vocal expression of their gender.

As noted earlier, certain social contexts – such as the presence of same-sex peers may trigger gender-typed behaviour (Banerjee & Lintern, 2001). The present study raises the question of whether the control of acoustic parameters as reported in this study contributes to this self-presentation of gender. Several studies (Biernat, 1991; O'Brien & Huston, 1985; Serbin et al., 2001) have found that Western children acquire gender stereotypes in behaviour and appearance by three years of age (and increase their gender-typed associations as they get older), but to our knowledge no research has focused on the acquisition and role of voice stereotypes in children. The development of voice control in the expression of gender in children's everyday speech therefore remains to be studied. Moreover, given the importance of social environment on children's gender identity, future studies should examine the role of parental-child interactions, peer interactions and child-directed media (e.g. advertising, cartoons) on voice gender acquisition and development in a range of cultures and societies.

Acknowledgements

We extend our thanks to the head teachers and staff at Hurst Pre-Prep School, Hurstpierpoint, Lewes YMCA, Lewes, Middle St Primary School, Brighton, and Holy Trinity CE Primary School, Horsham, who took the time and trouble to let us collect data in the pilot and final stages of this project. We are also grateful to the children and their parents who agreed to take part in the study.

References

Baker, S., Weinrich, B., Bevington, M., Schroth, K., & Schroeder, E. (2008). The effect of task type on fundamental frequency in children. *International Journal of Pediatric Otorhinolaryngology*, *72*(6), 885–889.

Banerjee, R., & Lintern, V. (2001). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408.

Bennett, S. (1983). A 3-year longitudinal study of school-aged children's fundamental frequencies. *Journal of Speech and Hearing Research*, *26*(1), 137.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2.17)[Computer program]. Retrieved January 1, 2012.

Busby, P. A., & Plant, G. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*, *97*, 2603.

Carnevale, V., Romagnoli, E., Cipriani, C., Del Fiacco, R., Piemonte, S., Pepe, J., Scillitani, A., et al. (2010). Sex hormones and bone health in males. *Archives of Biochemistry and Biophysics*, *503*(1), 110–117.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PloS one*, *7*(2), e31353.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*, 1511.

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*(1), 30–57.

Hay, D. F., Nash, A., Caplan, M., Swartzentruber, J., Ishikawa, F., & Vespo, J. E. (2011). The emergence of gender differences in physical aggression in the context of conflict between young peers. *British Journal of Developmental Psychology*, *29*(2), 158-175.

Hewlett, N., & Beck, J. M. (2006). *An Introduction To The Science Of Phonetics*. Routledge.

Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental Trajectories of Sex Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years. *Child Development*, *79*(5), 1583–1593.

Kahane, J. C. (1978). A morphological study of the human pre-pubertal and pubertal larynx. *American Journal of Anatomy*, *151*(1), 11–19.

Karlsson, I. (1989). Dynamic voice source parameters in a female voice. *Speech Transmission Laboratory-Quarterly Progress and Status Report, Royal Technical Institute, Stockholm*, *1*, 75–77.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Leaper, C. & Smith, T. A. (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, *40*, 993-1027.

Leman, P. J., Ahmed, S. & Ozarow, L. (2005). Gender, gender relations, and the social dynamics of children's conversations. *Developmental Psychology*, *41*, 64-74.

Leman, P. J., & Tenenbaum, H. R. (2011). Practising gender: Children's relationships and the development of gendered behaviour and beliefs. *British Journal of Developmental Psychology*, *29*(2), 153-157.

Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's gender-based reasoning about toys. *Child Development*, *66*(5), 1453–1471.

Munroe, R. L., & Romney, A. K. (2006). Gender and Age Differences in Same-Sex Aggregation and Social Behavior A Four-Culture Study. *Journal of Cross-Cultural Psychology*, *37*(1), 3–19.

O'Brien, M., & Huston, A. C. (1985). Development of sex-typed play behavior in toddlers. *Developmental Psychology*, *21*(5), 866–871.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology Vol. 3: Social, Emotional, and Personality Development* (6th ed., pp. 858-932). Hoboken, NJ: Wiley.

Sachs, J., Lieberman, P. & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In: Shuy R. W. & Fasold R. W., Editors (Eds.). *Language attitudes: Current trends and prospects* (74–84). Washington, DV: Georgetown University Press.

Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, *25*(1), 7–15.

Stevens, K. N. (2000). *Acoustic Phonetics*. MIT Press.

Titze, I. R. (1994). *Principles of voice production* (pp. 279-306). Englewood Cliffs: Prentice Hall.

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1510.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., Chung, M. K. (2011). Developmental Sexual Dimorphism of the Oral and Pharyngeal Portions of the Vocal Tract: An Imaging Study. *Journal of Speech, Language, and Hearing Research*, *54*(4), 995.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, *110*, 464.

Zosuls, K. M., Martin, C. L., Ruble, D. N., Miller, C. F., Gaertner, B. M., England, D. E., & Hill, A. P. (2011). 'It's not that we hate you': Understanding children's gender attitudes and expectancies about peer relationships. *British Journal of Developmental Psychology*, *29*(2), 288-304.

Appendix A - Acoustic expression of gender in the human voice

The voice can be seen as an important dimension of gender identity. Two key acoustic cues that are known to convey gender information in the voice are its fundamental frequency (F0, equal to the rate of vocal fold vibration), and its formant frequencies (Fi, the vocal tract resonances resulting from the filtering action of vocal tract cavities as the voiced sound travels from the glottis to the mouth (Titze, 1994)). Adult men typically have a lower F0 than women, resulting in lower perceived pitch, as well as lower formant values, resulting in a more baritone voice (Fitch & Giedd, 1999). In pre-pubertal children, while sexes do not differ in F0 (Baker et al., 2008; Lee et al., 1999), boys typically speak with lower formant values than girls (Bennett, 1983; Busby & Plant, 1995; Sachs et al., 1973), although the extent of these differences is both vowel- and formant-dependent (Whiteside, 2001). Moreover, adult listeners can identify the gender from the voice alone of children as young as four years (Perry et al., 2001) with a good level of accuracy (from 66% to 81% (Karlsson, 1989; Sachs et al., 1973)), suggesting that these acoustic differences in pre-pubertal children's speech are perceptually relevant.

In adults the observed differences in F0 and formants are largely due to testosterone-driven changes to the vocal apparatus occurring at puberty (Titze, 1994). During this period, males develop longer vocal folds than females, leading to a two-fold drop in F0, which is inversely proportional to vocal fold length. In addition, men develop longer vocal tracts than females, due to the male-specific secondary descent of the larynx (Fitch & Giedd, 1999; Vorperian, 2007) and increased height (Carnevale et al., 2010). As vocal tract resonances (formants) and their overall spacing are inversely related to vocal tract length, men's longer tracts are characterised by lower formant values and narrower spacing than women (Titze, 1994).

However, while the sex dimorphism of the adult voice is mainly determined by the underlying anatomical dimorphism, morphometric studies of the pre-pubertal vocal apparatus have failed to identify substantial sex differences that account for the observed sex differences in their formant frequencies (Fitch & Giedd, 1999; Vorperian et al., 2007; Vorperian et al., 2011). This has led to the hypothesis that the vocal expression of gender in children may involve children learning gender-related articulatory strategies, so that they "sound" like a male or a female by following some aspects of the adult voice's dimorphism. For example, boys may protrude their lips to lengthen their vocal tract, thus lowering their formants and narrowing formant spacing, while girls may spread their lips to achieve the opposite effect (Sachs et al., 1973). However, this hypothesis has never been investigated.

Appendix B - Acoustic Analysis Details

The speech material consisted of nine non-diphthong vowels of British English embedded in CVC words (/ae/ "hat", /eh/ "bed", /er/ "bird", /iy/ "feet", /ih/ "pig", /ah/ "duck", /aa/ "box", /uh/ "book", /uy/ "boot"). Prior to the analysis, each sample was renamed with a random identifier in order to ensure blind testing.

Fundamental Frequency

For the F0 analysis, a custom script was written in *PRAAT v.5.2.17* (Boersma & Weenink, 2011). The script utilises the PRAAT autocorrelation algorithm "to Pitch" to estimate the F0 contour. Each sample was processed using pitch floor 60Hz and ceiling 500Hz, time step 0.01s. The resulting F0 contour was double checked by visual inspection of the sample spectrogram during processing, and erroneous estimates were manually corrected. From the contour the script calculates mean F0 (F0), standard deviation (F0SD) and coefficient of variation (F0CV). F0SD gives an indication of absolute variation, but does not account for the logarithmic relationship between absolute and perceived pitch. For example, a 200Hz difference between 200Hz and 400Hz will be perceived as greater than between 400Hz and 600Hz because pitch is based on the ratio of the two frequencies (2 and 1.5) rather than the absolute difference. Therefore we also included F0CV, as this measure is calculated relative to F0 magnitude (F0SD/F0), and thus is independent from F0, reflecting the perceptual scaling of F0 variation (Gaudio, 1994; Lee et al., 1999). Moreover, in order to account for the logarithmic perception of F0 (Stevens, 2000), shifts in F0 were reported in semitones (number of semitones (ST) = 39.863 x log($F0_2$/$F0_1$ (Hewlett & Beck, 2006)) as well as in Hz.

Formant Frequencies

The first four formants (F1-F4) of each vowel were tracked automatically using PRAAT's Linear Predictive Coding "Burg" algorithm. The parameters for formant analysis were set as: number of formants 5, max formant 6000–6600 Hz, and dynamic range 30dB. The length of the analysis window was 0.025s. The accuracy of each formant track was manually checked and the script parameters changed to align the tracks with the formants shown in the sample spectrogram. Evaluation of each formant centre frequency was then measured from the central, steady state portion of each vowel.

Formant Spacing

The difference between any two adjacent formant frequencies, also defined as formant spacing ((1) $\Delta F = F_{i+1} - F_i$), was measured by using the model in Cartei and colleagues (2012), which is described in Reby and McComb (2003). The model approximated the vocal tract to a quarter-wave length resonator of uniform cross-sectional area. Under such model, each formant can be estimated by the following formula:

(2) $Fi = \frac{(2i-1)c}{4VTL}$

where *i* is the formant number, *c* is the speed of sound in a mammal vocal tract (350m/s), *VTL* is vocal tract length (in m) and *F*i is the frequency (in Hz) of *i*th formant. From formulas 1 and 2, it follows that ΔF can be calculated as the slope of a regression model (formula 3) with the observed Fi values (y-axis) plotted against the expected formant positions (x-axis):

(3) $Fi = \frac{(2i-1)}{2}\Delta F$

While individual formants are sensitive to deviations from the model due to the non-uniform vocal tract shapes required to express the different sounds, the formant spacing is an average of adjacent formant differences, and thus provides an overall estimate of spectral dispersion, which is less sensitive to such deviations. Thus, the spacing between the resonant frequencies will decrease as the vocal tract length increases, and will increase as the vocal tract length decreases.

Appendix C – Supplementary Tables

Table 4.4.3

Mean and standard deviation (SD) in Hz for the acoustic parameters of boys in the masculinised, natural and feminised conditions

Acoustic Parameters	Masculinised		Natural		Feminised	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
F0	243.5	32	249.6	29	307.2	62
F0SD	27.8	14	25.2	12	34.1	17
F0CV	0.12	0.06	0.11	0.06	0.12	0.07
F1	698	68	710	55	735	75
F2	1997	111	2046	125	2123	168
F3	3337	221	3375	227	3511	250
F4	4366	232	4487	236	4615	262
ΔF	1284	69	1313	68	1355	80

Note. Acoustic parameters: Fundamental Frequency (F0), F0 Standard Deviation (F0SD), Coefficient of Variation (F0CV), Individual values for the first four formants (F1-F4), Formant Spacing (ΔF)

Table 4.4.4

Mean and standard deviation (SD) in Hz for the acoustic parameters of girls in the masculinised, natural and feminised conditions

Acoustic Parameters	Masculinised		Natural		Feminised	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
F0	234.6	30	249.1	26	270.2	50
F0SD	36.1	42	28.0	16	26.0	15
F0CV	0.15	0.14	0.12	0.07	0.10	0.05
F1	685	77	753	60	779	64
F2	2072	114	2207	112	2254	120
F3	3324	229	3467	163	3574	156
F4	4446	228	4602	162	4703	203
ΔF	1301	67	1355	46	1389	53

Note. Acoustic parameters: Fundamental Frequency (F0), F0 Standard Deviation (F0SD), Coefficient of Variation (F0CV), Individual values for the first four formants (F1-F4), Formant Spacing (ΔF)

Study 4.1: Self-awareness of Voice Gestures in Pre-pubertal Children

Note. This section is not part of the published manuscript.

As part of the experimental procedure in Study 4, I also investigated children's awareness of the contribution of F0, formant shifts and related articulatory gestures (lip/laryngeal movements) in masculinising or feminising their voices.

Procedure

After the recordings, all children were asked to spontaneously describe what they did to sound "like a boy" or "like a girl" as much as possible. Then, I asked them whether they noticed any changes in pitch, and in lip spreading, rounding or protruding (I showed these movements to them by moving my lips). They were also asked whether they noticed any vertical movement of the larynx (I indicated the laryngeal notch by placing my fingers on my throat), when imitating boys' and girls' voices.

Results

When asked to spontaneously describe their strategies to alter their voice gender, 14 (out of 19) girls reported that to sound "like a boy" they made their voices lower, compared to five (out of 15) boys, χ^2 (34) = 5.54, p = .019. This was the only significant association between sex and type of strategy. Additionally, three boys and two girls said that they deepened their voices, χ^2(34) = .60, p = .439, while one boy and one girl reported tilting their head down, χ^2(34) = .030, p = .863. To sound "like a girl", nine boys and 12 girls spontaneously reported making their voices higher, χ^2(34) = .035, p = .851. Additionally, one boy reported making his voice sound "dollier", χ^2(34) = 1.31, p = .253, while one girl reported tilting her head up, χ^2(34) = .813, p = .367. When given a choice of possible gestures, two boys and one girl mentioned lowering their pitch to masculinise their voices, χ^2(34) = .679, p = .410, and two boys and one girl noticed raising their pitch to feminise them, χ^2 (34) = .679, p = .410, while the rest of the children reported not knowing what voice pitch was. Additionally, none of the children reported any lip or laryngeal movements underlying their voice gestures.

Discussion

Overall, children's descriptions reveal some awareness of the perceptual outcome of their gestures: the majority of girls (but only about one third of the boys) described lowering their voices to masculinise their voices, and most children reported raising their voices to feminise them. Children, however, were not able to describe their adjustments in terms of pitch or vocal tract adjustments (by lip or laryngeal movements). Interestingly though, one boy and one girl reported tilting their head down to sound "like a boy", while one girl reported tilting her head up to sound "like a girl". Indeed, head tilting can masculinise or feminise one's voice: a low head position, for example, causes the larynx to drop, thus lengthening the vocal tract (decreasing formant spacing), and potentially lowering F0 due to the rotation of the cricoid cartilage along the cervical lordosis, which decreases vocal fold tension, overcoming the associated shortening of the vibrating folds (Honda et al., 1999). Vice versa, tilting the head up shortens the tract, thus increasing formant spacing, as well as vocal fold tension, thus raising F0 (Honda et al., 1999). In contrast to children, Cartei and colleagues (2012) reported that all men and women described their masculinised voices as "deeper" or "lower" than they normal speaking voices, and their feminised voices as "higher" or "softer". Additionally, most adults of both sexes reported being aware of pitch adjustments, and of vocal tract adjustments (especially larynx lowering by men). Age-related differences in participants' self-reports of voice gestures may reflect individuals' increased knowledge of voice gender differences, especially as these become more marked due to the anatomical changes occurring during male puberty, but may also reflect developmental sex-specific processes in terms of gender identity and stereotyping (Banerjee & Lintern, 2000; Berk, 2000; Biernat et al., 1991; Miller et al., 2009; Serbin et al., 1993). The extent to which awareness of voice gestures and underlying articulatory behaviours correlate with their control is an exciting area of future research.

References

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control. *Language and Speech*, *42*(4), 401–411. doi:10.1177/00238309990420040301

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: a developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351–365.

Serbin, L. A., Powlishta, K. K., Gulko, J., Martin, C. L., & Lockheed, M. E. (1993). The Development of Sex Typing in Middle Childhood. *Monographs of the Society for Research in Child Development*, *58*(2), i. doi:10.2307/1166118

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408.

Berk, L. E. (2000). Development of Sex-related differences and gender roles. In: *Child Development : 5th Edition* (521-526). Allyn and Bacon.

Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of Gender Stereotype Domains: Developmental and Gender Differences in Children. *Sex Roles*, *60*(11-12), 870–881.

Study 5:

Spontaneous voice gender imitation abilities in adult speakers

Note. This study is published as: Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PloS ONE*, *7*(2), e31353.

Abstract

The frequency components of the human voice play a major role in signalling the gender of the speaker. A voice imitation study was conducted to investigate individuals' ability to make behavioural adjustments to fundamental frequency (F0), and formants (Fi) in order to manipulate their expression of voice gender.

Thirty-two native British English adult speakers were asked to read out loud different types of text (words, sentence, passage) using their normal voice and then while sounding as 'masculine' and 'feminine' as possible. Overall, the results show that both men and women raised their F0 and Fi when feminising their voice, and lowered their F0 and Fi when masculinising their voice.

These observations suggest that adult speakers are capable of spontaneous glottal and vocal tract length adjustments to express masculinity and femininity in their voice. These results point to a "gender code", where speakers make a conventionalised use of the existing sex dimorphism to vary the expression of their gender and gender-related attributes.

Introduction

The human voice is highly sexually dimorphic. Alongside other properties that distinguish male from female voices, such as intonation (McConnell-Ginet, 1978), duration (Ericsdotter & Ericsson, 2001; Simpson, 2003) and speech rate (Byrd, 1992; Whiteside, 1996), the main cues to speaker gender are fundamental frequency (F0 or its perceptual correlate "pitch") and formant frequencies (Fi, mainly responsible for the perception of "timbre"), which together account for 98.8% of the perceived voice dimorphism (Bachorowski & Owren, 1999).

These differences stem from the testosterone-driven enlargement of the larynx and the increase in the length of the vocal tract that accompany male puberty (Titze, 1994). During this time, the male larynx outgrows the female larynx by 40% (Titze, 1994), increasing vocal fold length by 60% on average (reaching 16mm in adult males, and 10mm in adult females (Hirano, Kurita & Nakashima, 1981)). As F0 is based on the rate of vocal fold vibration, which in turn is inversely proportional to the square root of the vocal fold tissue length, men's F0 (about 120Hz) becomes on average 80Hz lower than women's (about 200Hz) giving male speakers their characteristically lower-pitched voice (Titze, 1994). Between-sex differences in formant frequencies are related to differential body growth, with adult men being 7% taller than women on average (Gaulin & Boster, 1985) and to the male-specific second descent of the larynx, which together contribute to men's vocal tract being on average 18cm, compared to women's 15cm (Vorperian et al., 2009). Because formant frequencies are negatively correlated with the length of the vocal tract (Fant, 1960), male speakers produce lower Fi values and therefore a formant spacing (ΔF) that is about 15%–20% narrower than in female speakers (Fant, 1966; Goldstein, 1980), which results in male voices having a more "baritone" timbre (Fitch & Giedd, 1999).

Variation in gender expression, however, cannot be entirely determined by these hormonal and size-related sex differences in the vocal apparatus. For example, acoustic analyses of pre-pubertal children's voices consistently show that boys speak with lower formants than girls (Bennett, 1981; Lee, Potamianos & Narayanan, 1999; Perry, Ohde & Ashmead, 2001; Whiteside & Hodgson, 2000), while perceptual studies show that children's voice gender can be identified in children as young as four years old, despite the fact that the anatomy of the vocal apparatus does not significantly differ between the two sexes until the pubertal age (Fitch & Giedd, 1999; Perry et al., 2001; Vorperian & Kent, 2007). These observations suggest that children acquire (consciously or unconsciously) gender-specific articulatory behaviours during development, and that speakers develop a knowledge of how a "male" or a "female" should sound, with male voices being lower-pitched and "deeper", while female voices being higher-pitched and "lighter". These differences in formant frequencies also suggest a possible role for lip protrusion (or spreading) and larynx lowering (or raising) in vocal tract length adjustments during speech, as possible articulatory gestures used by speakers in order to masculinise or feminise their voices. Thus, on top of the static, bio-hormonally determined differences, our voice contains dynamic and behaviourally controlled acoustic cues (in particular F0 and formants) for the expression of gender and gender-related attributes. However, the nature and the extent of their role have not yet been systematically investigated.

Hypotheses

The current study explores the ability of adult speakers to alter the femininity and masculinity of their voices during an imitation experiment, as well as the extent to which they are aware of the nature of the underlying articulatory gestures that they use to make these alterations. We predict that both male and female speakers will lower their mean F0, reduce its variation, and lower their Fi, thus narrowing ΔF, when trying to sound as "masculine" as possible, whilst they will increase their mean F0 and its variation, as well as raise Fi, thus widening ΔF, to sound as "feminine" as possible. In addition, we hypothesise that speakers will round their lips in order to lengthen their vocal tract when masculinising their voice, and spread their lips to shorten their tract when feminising their voice. We also investigate male and female speakers' awareness of the contribution of F0, formant shifts, and related articulatory gestures (lip/laryngeal movements) to the vocal exaggeration of masculinity and femininity.

Materials and Methods

Subjects

Participants were 15 female and 17 male undergraduate students from the University of Sussex (UK), between 18 and 45 years of age (M = 22.56, SD = 6.4) with no self-reported history of speech, language, or hearing disorders. All were native speakers of British English. Informed written consent was obtained for all participants before study entry.

Procedure

Voice data were collected from individual speakers in a sound-attenuated booth at the University of Sussex. Participants were seated in a comfortable chair wearing a hat fixed to the chair in order to limit head movement, and were audio recorded with a high-fidelity microphone (*AKG Perception 220*).

Each participant was asked to read three different types of written stimuli out loud, first using their normal speaking voice (neutral condition), then sounding as 'feminine' as possible (feminine condition) and then as 'masculine' as possible (masculine condition), in alternate order. The material included a list of vowels embedded in a CVC context (vowel task), one short sentence that included many of the vowel sounds present in the vowel task (sentence task), and a 168 word passage comprised of several sentences (passage task). The order of presentation of the CVC words was randomised across participants to avoid serial order effects. Participants were allowed to progress at their own pace, choosing to continue to the next word only when ready. The word and sentence sequences were shown on a computer monitor, using a script written in PsyScope X Build 57. The text extract was shown in Microsoft Word 2007.

Participant's height and weight were measured prior to collecting the speech sample. Height measurements were recorded to the nearest 0.1cm, using a freestanding *Seca Leicester* stadiometer. Participants took their shoes off and stood with their shoulders flush to the stick and their heads level and oriented forward. Body weight was measured to the nearest 0.1kg using a *PS250* veterinary floor scale. Means, standard deviations, and range values for participants' body size measurements are reported in Table 4.5.1.

After completion of the vocal task, the experimenter went over a questionnaire with participants about the strategies they used to masculinise and feminise their voices, and recorded their responses on paper. The questionnaire began with a series of open questions, followed by multiple-choice questions on several vocal and articulatory gestures.

Table 4.5.1

Mean, standard deviation (SD) and range values of speakers' height and weight

	Mean	SD	Range
Men			
Height (cm)	181.9	6.0	171.0–188.0
Weight (Kg)	73.3	6.9	64.3–88.7
Women			
Height (cm)	163.3	7.1	149.6–173.6
Weight (Kg)	59.9	10.9	41.7–70.5

Visual Measurements

For each participant, we measured lip spreading (LS), the horizontal distance between the two mouth corners, and openness (LO), the vertical distance between the centres of the upper and lower lips. In order to take these measurements, the horizontal mouth corners and the upper and lower centre lips were marked using a black makeup pencil (horizontal lines for the upper and lower lips, vertical lines for the mouth corners). The lip ratio (LR) for each participant was also calculated as the ratio between his or her lip spreading and openness. Video recordings of the participants were taken using a *Sony HDR-TG3E handycam*. The visual measurements were taken from stills captured using Apple iMovie version 8.0.6 of the vowel task just after the participant had uttered the first consonant. Markers were then used to extract the horizontal (lip spreading) and vertical (lip openness) mouth distances using the line drawing function in Adobe Illustrator CS5.

Acoustic Measurements

The stimuli consisted of nine monophthong British vowels in /CVC/ sequences (had /æ/, head /e/, hud /ʌ/, heed /iː/, hid /ɪ/, heard /ɜː/, hod /ɒ/, hood /ʊ/, who'd /u/), the sentence “where were you a year ago?” and an extract from the “Rainbow Passage” (Fairbanks, 1960). A custom script was written in PRAAT v.5.0.3 (Boersma & Weenink, 2006) to process the collected audio samples. The script assigned a random identifier to each sample in order to ensure blind analysis. It then allowed the experimenter to set the analysis parameters and to visually compare the fundamental and formants frequencies against a broadband spectrogram. The analysis parameters were adjusted when the computed values departed from the visually estimated fundamental and formant frequencies.

Fundamental Frequency. For the F0 analysis, the script used the PRAAT autocorrelation algorithm “to Pitch”, which estimates the F0 contour, from which the script derived mean F0 (F0*mean*), F0 standard deviation (F0*SD*) and the coefficient of variation (F0CV). F0*CV*, which is given by F0*SD*/F0*mean*, provides a measure of the magnitude of F0 variation relative to the mean, which reflects the logarithmic perception of pitch and therefore is a better estimate of F0 variation than its absolute estimate given by F0*SD* (Lee et al., 1999). Perceptually, a voice with lower F0*CV* has a more monotone quality than a voice with higher F0*CV*. The parameters for F0 analysis were set as: pitch floor 30Hz and ceiling 500Hz for male speakers, 60Hz and 500Hz for female speakers, time step 0.01s.

Formant Frequencies. For formant (Fi) analysis, the script used PRAAT's Linear Predictive Coding "Burg" algorithm in order to estimate the formant centre frequencies for the first four formants (F1–F4). The parameters for formant analysis were set as: number of formants 5, max formant 5000 Hz for male speakers and 5500Hz for female speakers, and dynamic range 30dB. The length of the analysis window was 0.0025s in the vowel and sentence tasks, and 0.005s in the passage task.

Formant spacing. The centre frequencies for F1–F4 of each sample were used to calculate its average formant spacing (ΔF), which is the distance between any two adjacent formants:

(1) $\Delta F = F_{i+1} - F_i$

ΔF was calculated by forcing the observed Fi values to fit the vocal tract model described in the source-filter theory (Fant, 1960). In this model, the vocal tract has a uniform cross-sectional area along its entire length, which approximates the production of the vowel "schwa" (/ə/). Thus, the vocal tract acts as a quarter-wave resonator, closed at the glottis and open at the mouth, and the vocal tract resonances are given by:

(2) $F_i = \dfrac{(2i-1)c}{4VTL}$

where Fi is the i^{th}-formant, c is the speed of sound in the human vocal tract (approximated to 350 m/s) and VTL is the length of the resonator. From (1) and (2), it follows that individual formants are related to ΔF by:

(3) $F_i = \dfrac{(2i-1)}{2}\Delta F$

ΔF can therefore be calculated as the slope of the linear regression expressed in equation (3), by plotting the observed Fi (y-axis) against the expected 2i-1 formant positions (x-axis), and with the intercept set to 0 (Reby & McComb, 2003).

Whilst the specific variation of formants in vowels other than the "schwa" requires more complex models than the uniform quarter wavelength resonator used here

(Stevens, 2000), the average distribution of formants at suprasegmental level approaches a constant that corresponds to the *ΔF* predicted by such a model (Titze, 1994). The adequacy of this method is illustrated by estimations of *ΔF* based on published acoustic data (Appendix A). It is also consistent with perceptual observations: Smith and Patterson (2005) report that Δ*F* differences re-synthesised via linear compression/expansion of the vowel spectral envelope correlate strongly with listeners' cross-class judgments of speaker's age, sex and size (man, woman, boy, girl). More recently, Pisanski and Rendall (2011) also found that small (12% or 18%) uniform increments in Fi negatively correlate not only with the perceived size, but also with the masculinity of speakers within the same sex and age group.

Statistical Analyses

Two-way mixed ANOVAs were used to investigate the overall effect of sex (group factor) and condition (as a three-level repeated factor: neutral, masculine, feminine) on each of the acoustic parameters F0*mean*, F0*CV*, Fi and Δ*F*, and on the visual parameters LS, LO and lip ratio. We also tested for differences across conditions for male and female speakers separately, running separate one-way repeated ANOVAs within each sex with condition as the factor variable and using contrasts between neutral and masculine, and neutral and feminine conditions. Levene's tests were used to check for equality of variance, and the data were log-transformed when the assumption was violated. A Mauchly's test was applied in order to check sphericity, and sphericity violations were corrected for with the Greenhouse-Geisser ε. All statistical analyses were run using SPSS v.18.

Results

The results of the ANOVAs performed on the acoustic measures are presented in Table 4.5.2 (vowel task), Table 4.5.3 (sentence task), and Table 4.5.4 (passage task) in Appendix B. The means and standard deviations of the acoustic measures, and the F and p-values of the associated contrast are provided separately for female and male speakers in Tables 4.5.5, 4.5.6, 4.5.7 and 4.5.8, also in Appendix B.

Fundamental Frequency

There was a significant main effect of sex on F0*mean* in all three reading tasks, indicating that male speakers had a lower mean F0 than female speakers across conditions, in line with the well-established sexual dimorphism in mean F0 between the two sexes.

There was also a significant main effect of condition on F0 across the three tasks. Separate ANOVAs revealed that both male and female speakers significantly raised their F0 when feminising their voice and dropped their F0 when masculinising their voice (except when reading the passage, where the difference between neutral and masculine conditions was not significant). The largest drop in F0 between speakers' natural and masculinised voice occurred when reading the sentence, with male speakers significantly dropping their F0 by about 7% from 110.6Hz to 103.8Hz (Table 4.5.6) and female speakers by about 8% from 196.2Hz to 178.8Hz (Table 4.5.5). Both male and female speakers also significantly raised their F0 when feminising their voices. The largest change in F0 between speakers' natural and feminised voice occurred when reading the sentence, with male speakers raising their F0 to 162.2Hz (about 40% rise (Table 4.5.6)) and female speakers to 256.7Hz (about 24% (Table 4.5.5)), whereas the smallest, yet significant, rise was recorded in reading the passage, 28% for men (Table 4.5.6) and 20% for women (Table 4.5.5). The interaction effect between condition and sex was not significant.

Fundamental Frequency variation (F0CV)

The effect of sex on F0*CV* was not significant for vowels, but was significant in the other two tasks, indicating that, overall, men spoke with a narrower dynamic range than women.

There was also a significant main effect of condition in the sentence and passage, but not for the vowels. Contrasts revealed that male speakers' F0*CV* was not significantly lower when sounding as masculine as possible as when speaking normally (although a non-significant trend was observed for the passage, p = .096 (Table 4.5.8)). Female speakers' F0*CV* was significantly lower in the masculine condition, but only when reading the passage out loud (Table 4.5.7). There was a non-significant trend for male speakers to raise F0*CV* when reading the passage in a feminised voice, p = .060

(Table 4.5.8), while female speakers significantly increased their F0*CV* to feminise their voice only in the vowel task (Table 4.5.7).

Formant frequencies

There was a significant main effect of sex on Fi in all three reading tasks indicating that male speakers' formants were lower than female speakers' across conditions.

There was also a significant main effect of condition on Fi across the three tasks. Contrasts revealed that, when asked to sound as masculine as possible, men lowered all their formants, except for F1 across conditions, F2 and F3 in the sentence task, for which no significant differences were found (Table 4.5.8). Female speakers also significantly lowered their formants when sounding as masculine as possible for all three tasks, except for F1 in the sentence task (Table 4.5.7).

When asked to sound as feminine as possible, male speakers significantly raised their formants, except for F1 across conditions and F2 in the sentence task (Table 4.5.8). Females also showed an overall tendency to raise their formants, although statistical significance was only reached for F4 in the vowel task, and F1, F2 and F4 in the sentence task (Table 4.5.7).

Linear mixed models testing for differences in Fi were run separately for each sex as a function of condition and vowel. The results are shown graphically in Figure 4.5.1. For both men and women, there were main effects of condition and vowel on each individual formant frequency, while no significant interaction effect between condition and vowel was found on Fi (see Table 4.5.9 in Appendix B).

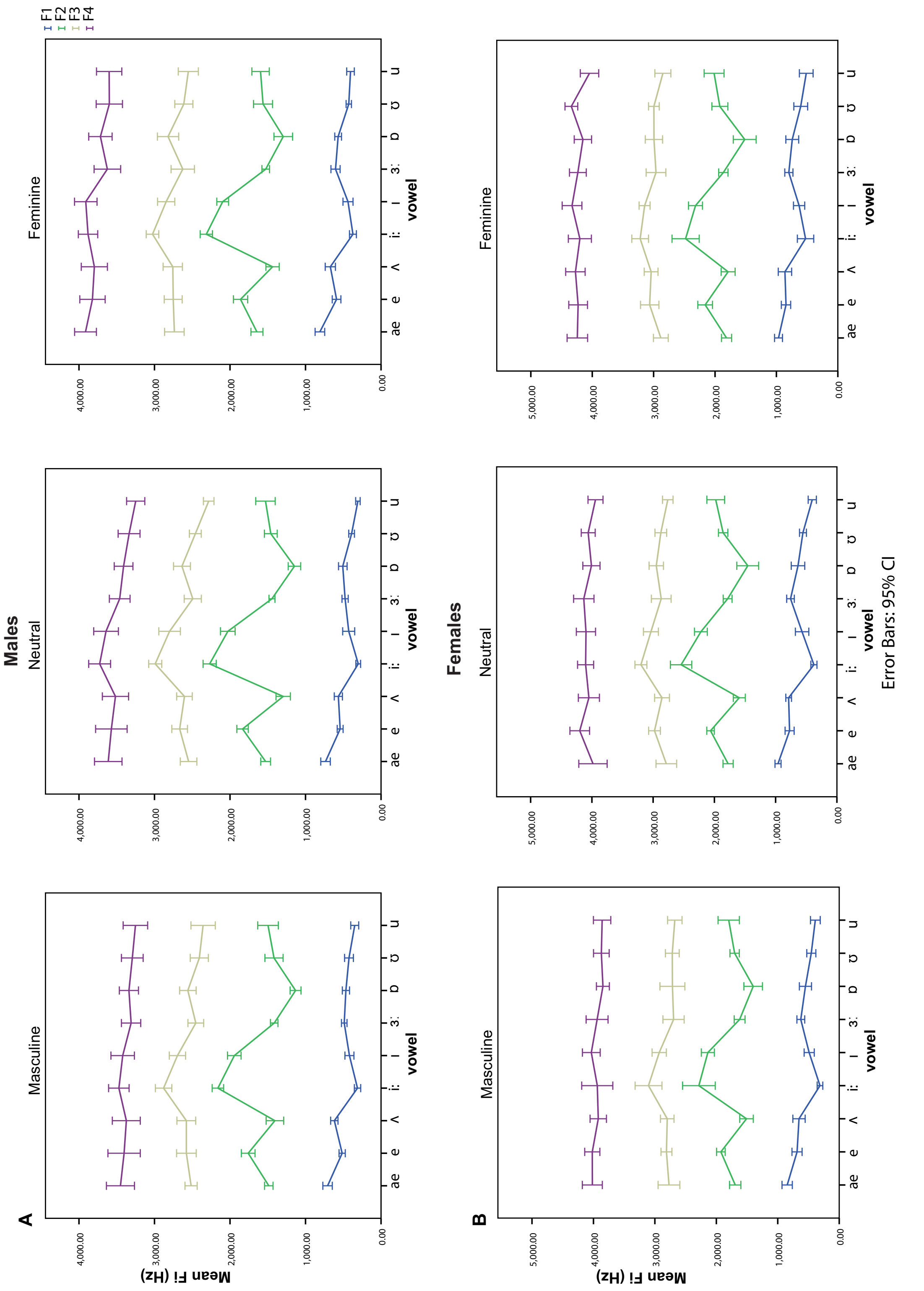

Figure 4.5.1. Formant values across vowels within each condition for male (A) and female (B) speakers.

The vowel spaces (Figure 4.5.2) show that the vowels in the neutral condition match the typical vowel distribution in F1/F2 space for both sexes, whilst the vowel spaces in the masculine and feminine conditions match the neutral vowel space in shape, but are smaller and globally shifted downward and left, and bigger and globally shifted upward and right, respectively.

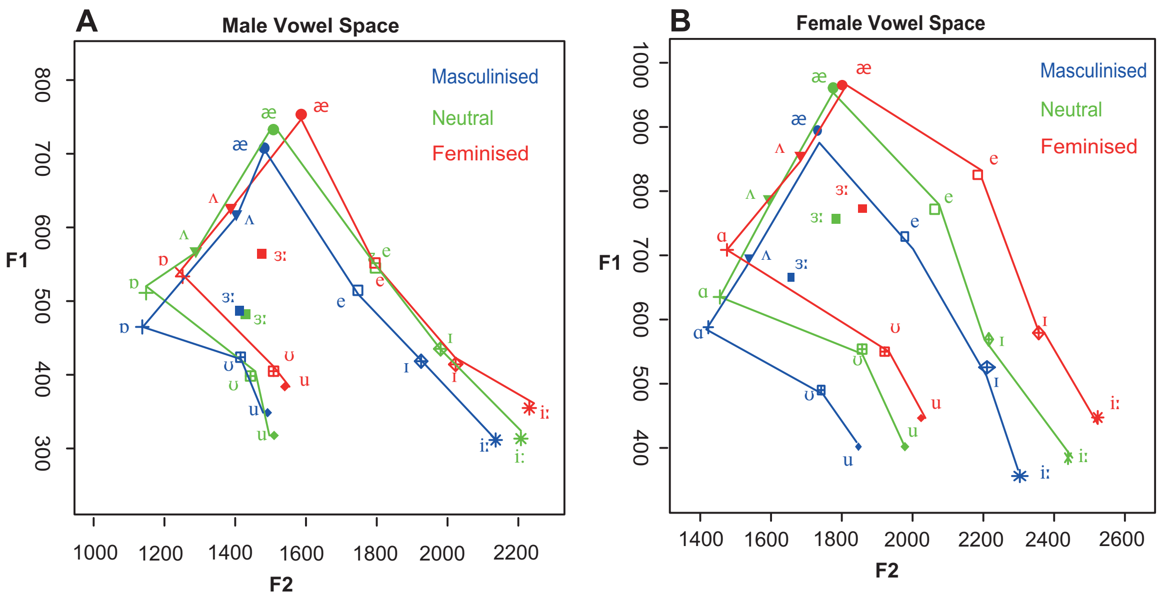

Figure 4.5.2. Vowel spaces of male and female speakers. Scatter plots of the mean frequency of F1 and F2 for the nine vowels spoken by men (A) and women (B) across the masculine, neutral, and feminine conditions. The overall vowel spaces are outlined by joining the isolated vowels through straight lines.

Formant spacing

There was a significant main effect of sex on *ΔF* in all the three reading tasks, indicating that male speakers had a narrower overall formant spacing (*ΔF*) than female speakers. There was also a significant main effect of condition on *ΔF* across the three tasks. The interaction effect between condition and sex was not significant. Contrasts revealed that both male and female speakers significantly narrowed their *ΔF* when masculinising their voice (Tables 4.5.8 and 4.5.7). In male speakers, the extent of this decrease varied from about 2% in the passage to 3% in the other two tasks (Table 4.5.6), while in female speakers it varied from about 3% in the passage to 5% in the other two tasks (Table 4.5.5). Male speakers also significantly widened their *ΔF* when feminising their voice (Tables 4.5.8), and the extent of this increase ranged from 3% in the passage to 6% and 5% in the sentence and vowel tasks (Table 4.5.6), respectively, while female speakers (Tables 4.5.5) increased their *ΔF* from 1% (passage, vowels) to 3% (sentence), reaching significance only in the sentence task (Tables 4.5.7).

Lip measurements

The mean and standard deviations for the lip measurements (in pixels) taken from the vowel task in each condition are presented in Table 4.5.10 (Appendix B). The main effect of sex was significant on Lip Spreading (*LS*), $F(1, 21) = 8.77$, $p = .007$, with women having a larger *LS* overall than men. There was also a significant main effect of condition on *LS*, $F(2, 42) = 13.86$, $p < .001$. Contrasts revealed that both men and women significantly reduced their LS when trying to sound as masculine as possible, and increased it when sounding as feminine as possible, albeit not significantly. No significant interaction between sex and condition was found, $F(2, 42) = 1.39$, $p > .05$.

There was a main effect of sex on Lip Openness (*LO*), $F(1, 21) = 7.95$, $p = .01$, which was greater in women than in men. The main effect of condition on *LO*, $F(2, 42) = 2.08$, $p > .05$, and the interaction effect of sex and condition, $F(2, 42) = 1.75$, $p > .05$, were not significant.

As for Lip Ratio (*LR*), the main effects of sex $F(1, 21) = 0.55$, $p > .05$, condition, $F(2, 42) = 2.2$, $p > .05$, and the interaction effect of condition and sex, $F(2, 42) = 3.71$, $p > .05$, were all not significant.

Moreover, separate mixed model tests of differences in all three parameters were run as a function of sex, condition, and vowel. There was a main effect of vowel on all three parameters (*LS*: $F(8, 535.02) = 36.35$, $p < .001$, *LO*: $F(8, 535.17) = 57.49$, $p < .001$, *LR*: $F(8, 535.41) = 24.26$, $p < .001$). The front vowels /æ/, /iː/, /ɪ/, showed the highest degree of lip spreading, while lowest degree of lip spreading was recorded for the back vowels /ɒ/, /ʊ/, /u/. High vowels /ʊ/, /u/ also showed the least degree of lip opening, whilst low vowels exhibited the greatest lip opening. The lip ratio was smallest for vowels /æ/, /e/. There were no interaction effects between condition and vowel, and sex and vowel, indicating that both men and women moved their lips in a similar way across all three conditions.

Participants' self-descriptions of vocal and articulatory gestures

Out of 17 male and 15 female speakers, when asked to spontaneously describe the strategies used to masculinise their voices, nine males and seven females replied that they made their voices sound deeper, $\chi^2(32) = .13$, $p = .723$, and eight males and four females said that they made them lower, $\chi^2(32) = 1.41$, $p = .234$. To feminise their voices, 12 males and seven females said that they made their voices higher, $\chi^2(32) =$

1.89, p = 1.69, and five males and four females reported making it softer, $\chi^2(32) = 0.30$, p = .86.

When given a choice of possible gestures, most participants reported changes in pitch: all 17 males and 14 females said that they lowered their pitch to sound more masculine, $\chi^2(32) = 1.17$, p = .279, and 16 males and 13 females said they raised their pitch to sound more feminine, $\chi^2(32) = 1.31$, p = .225. The majority of males also reported vocal tract length adjustments: 13 males reported the descent of their Adam's apple as a gesture to masculinise their voice, compared to six females, $\chi^2(32) = 4.39$, p = .036. This was the only significant association between sex and type of strategy. Six males also reported moving their Adam's apple up to feminise their voices, compared to four females, $\chi^2(32) = 2.76$, p = .599. As for lip movements, eight males and 11 females said they rounded their lips to sound more masculine, $\chi^2(32) = 2.28$, p = .131, while eight males and eight females said they spread their lips to sound more feminine, $\chi^2(32) = 1.25$, p = .723.

Discussion

We found that when untrained adult speakers were asked to sound as masculine or as feminine as possible, they altered the frequency components of their voice (F0 and formant parameters) by adjusting the rate of vibration of their vocal folds and by changing the apparent length of their vocal tract. This shows that adult speakers have some knowledge of the sexually dimorphic acoustic cues underlying the expression of gender in speech, and are capable of controlling them to modulate gender-related attributes. Below we discuss each F0 and formant parameter individually, focusing on their acoustic and perceptual relevance in relation to previous research. Then, we compare the observed manipulations to those used to express size, and, following the "size code" theory (Ohala, 1984), propose that a substantial proportion of gender-related vocal diversity in the human voice follows a "gender code", with speakers using learned vocal gestures to manipulate their voice gender. We also look at the interplay between the observed vocal tract adjustments (e.g. lip movements) and facial gestures in the context of gender expression. Finally, we propose some directions for future research.

Fundamental Frequency

For both sexes, the mean F0 measured in the neutral condition was comparable to previously reported F0 values in British English (Graddol & Swann, 1983). The observed sex dimorphism for this parameter (1.8) is in line with previous acoustic observations (Graddol & Swann, 1983) and can be mostly accounted for by the dimorphism in vocal fold length (1.6 (Titze, 1994)). The remaining 20% of dimorphism has been attributed to sex differences in vocal fold physiology (Titze, 1994), but may also point to differences in phonation behaviour (Rendall, Kollias, Ney & Lloyd, 2005; Simpson, 2009).

In both sexes, speakers lowered their F0 when masculinising their voices, and raised their F0 when feminising their voices, although in both conditions F0 remained within the expected range of their sex (around 100–160Hz for men, 170–260Hz for women (Hengton, 1989)). The F0 drop between the neutral and masculine conditions was about three times smaller than the F0 rise from the neutral to the feminine condition, with the smallest and non-significant drop being recorded for the passage. This could be a consequence of physiological constraints that make it more difficult for speakers to sustainably lower F0. Indeed, adult speakers speak with a mean F0 at the lower end of their physically attainable range in several languages (Traunmüller & Eriksson, 1994), and this is particularly the case of male speakers of British English (Graddol & Swann, 1983).

Perceptual studies with re-synthesised stimuli have previously reported that a F0 difference of 12% (Pisanski & Rendall, 2011; Puts, 2005) corresponding to twice the frequency discrimination threshold (or just-noticeable difference, JND) is required in order to elicit consistent results in masculinity discrimination performance. The observed differences in F0s between feminine/neutral and masculine/feminine conditions are above this threshold (Table 4.5.7, Table 4.5.8), suggesting that these differences are perceptually relevant. Psychoacoustic studies using natural stimuli, such as the ones produced here, could confirm whether this is the case and explore the perceptual relevance of the naturally occurring acoustic variation in the vocal expression of masculinity (or femininity).

F0 variation (F0*CV*) was higher for female speakers than for male speakers in reading the sentence and the passage; these longer stimuli may enable speakers to display more intonation variation (Thorsen, 1980). This result suggests that women speak with a wider dynamic voice range than men, which is in line with gender stereotypes (Henton, 1995), but contrasts with acoustic research adopting similar log scale conversions (Henton, 1989, 1995; Linke, 1973). In a comprehensive review of 40 years of research, Henton (1989) found that previously reported male-female differences in pitch range disappeared or were reversed when re-examined using the semitonal scale (semitones = 39.86 x log (F0max/F0min)). The discrepancy between the present results and Henton's may arise from the different methodologies used to model pitch perception. Although previous studies have cast doubts on the use of semitone scale as the most accurate measurement for F0 variation (Hermes & Van Gestel, 1991; Rietveld & Gussenhoven, 1985) the relative value of one method over the other is yet to be established.

Men exhibited a non-significant trend in increasing (decreasing) their F0*CV* when reading the passage to feminise (masculinise) their voices, but not in the other tasks. Women significantly increased their F0*CV* to feminise their voice when reading words, and decreased it to sound as masculine as possible when reading the passage. Although these differences are not consistent across all types of stimuli and between conditions, they nevertheless provide some indication that speakers may attribute wider intonation to female speech than male's, despite the fact that such attributions are largely unsupported by the literature (Hengton, 1989). Indeed, perceptual studies indicate that female speech is typically perceived as more 'melodious' than male's, both in pre-pubertal children's (Günzburger, Bresser & Keurs, 1987) and adults' voices (Kramer, Thorne & Henley, 1978). Greater F0 variation also elicits higher femininity ratings, while more monotonous voices are judged to be more masculine (Wolfe, Ratusnik, Smith & Northrop, 1990).

Formant frequencies and spacing

For both sexes, mean formant frequency values for the first four formants (F1–F4) in the neutral condition are within the range previously reported for adult speakers of Southern British English (Deterding, 1997; Hawkins & Midgley, 2005; Harrington, Kleber & Reubold, 2008), with the greatest percentage difference for F1 and the smallest for F3 (F1:22.2%, F2:13.3%, F3:11.1%, F4:13.6%) between the two sexes. A similar formant scaling dimorphism was found in a study of American English (Hillenbrand, Getty, Clark & Wheeler, 1995), although their scale factors do not entirely match the present results (F1: 18%, F2: 17%, F3: 14%).

Overall, speakers lowered their F1–F4 formants when asked to sound as masculine as possible and raised them to sound as feminine as possible. These global adjustments of formant frequency values are also reflected in the size and shifts of speakers' vowel spaces. Women's vowel space was larger and shifted top right relative to men's across conditions, in line with the known sex dimorphism (Rendall et al., 2005). However, both men and women's vowel spaces were larger, shifted upward to the right for the feminine condition, and were smaller and shifted downward to the left (Figure 4.5.2) in the masculine condition, compared to the neutral condition. This indicates that speakers exaggerated speech patterns typical of the two sexes in order to masculinise and feminise their voices.

Formant spacing (ΔF) values in the neutral condition were also comparable to those reported in the literature for both adult men (1005 Hz (Feinberg, Jones, Little, Burt & Perrett, 2005)); 991Hz, as calculated from F1–F4 values (Pisanski & Rendall, 2011)) and women (1167 Hz (Pisanski & Rendall, 2011)). Moreover, men's ΔF was on average 15% lower than women's, in line with the ΔF dimorphism reported in previous studies (Peterson & Barney, 1952; Pisanski & Rendall, 2011) and comparable to the 15%–20% baseline difference in anatomical vocal-tract length between the two sexes (Fant, 1966; Goldstein, 1980).

Consistent with our predictions, speakers widened their ΔF to feminise their voices and narrowed it to masculinise them, with wider shifts in formant values being observed when imitating opposite gender attributes than when exaggerating their own gender: averaged across reading tasks, men narrowed their ΔF by 2.7% to masculinise their voices, whilst women widened it by 1.9% to feminise theirs, whereas men widened their ΔF by 5.5% to feminise their voices and women narrowed it by 4.3% to masculinise theirs. These ΔF differences in the expression of gender-related attributes typical of the opposite sex correspond to the limit between the male upper and female lower ΔF ranges (Smith & Patterson, 2005).

Perceptually, the ΔF differences observed here between the natural and experimental conditions as well as between feminised and masculinised conditions (see

Tables 4.5.7, 4.5.8) are less than one JND (about 6%) for ΔF (Rendall et al., 2005). Thus, in combination with the percentage differences on F0 reported above, our study indicates that, although speakers adjust both F0 and ΔF to express gender-related attributes, only the F0 adjustments are likely to be perceived. Ultimately, by manipulating ΔF while preserving F0 and vice versa, future studies could look at the perceptual discriminability and relative salience of these two parameters in listeners' voice-based judgments of speakers' masculinity and femininity.

Is there a gender code?

Indications that adjustments in F0 and Fi parameters comparable to those observed in this study play a role in the expression of voice gender and related attributes are widespread in the literature on the sex dimorphism in the human voice. Despite having virtually the same vocal anatomy, pre-pubertal boys speak with lower formants than girls (Busby & Plant, 1995; Lee et al., 1999; Sachs, Lieberman & Erickson, 1973; Vorperian & Kent, 2007), suggesting that children acquire sex-specific behaviours, such as vocal tract gestures involving lip movements, to express their gender (Sachs et al., 1973). Acoustic studies of adult speakers also report within-sex differences in F0 and Fi that cannot be solely explained by anatomical differences. For example, in a cross-cultural study, Majewski and colleagues (1972) found that American men speak with a lower pitch (118.9Hz on average) than their Polish counterparts (137.6Hz on average), while Ohara (2001) found that Japanese women raise their pitch when speaking in their native language and lower it when speaking in English, in line with femininity definitions in Japanese society. Additionally, research on the vocal expression of sexual orientation shows that, while homosexual speakers' voices do not differ in mean F0 from their heterosexual counterparts (Gaudio, 1994; Rendall, Vasey & McKenzie, 2008) they display a partial shift of formant values towards those typical of the opposite sex (Munson, McDonald, DeBoe & White, 2006; Pierrehumbert, Bent, Munson, Bradlow & Bailey, 2004), even after controlling for body size (Rendall et al., 2008). Several perceptual studies also report that listeners rate adult voices characterised by higher pitch and formant values as more "feminine" (Pierrehumbert et al., 2004; Collins & Missing, 2003), while speakers with lower pitch and formant values are rated as more "masculine" (Hillenbrand et al., 1995; Munson & Babel, 2006; Rendall et al., 2005).

These observations suggest that speakers spontaneously use a “gender code”, making a conventionalised use of the existing sex dimorphism in the frequency components of their voice to vary the expression of gender and related (e.g. masculinity/femininity) characteristics. We draw a parallel between this gender code and Ohala’s (1984) “ size code” hypothesis, in which animal callers are expected to exploit the inverse correlation between resonator size and its resulting frequency in order to encode size and related (e.g. dominance/submission) attributes. Human male speakers have been shown to lower (or rise) F0 and Fi when they perceive themselves to be more (or less) dominant than their interlocutors (Puts, Hodges, Cardenas & Gaulin, 2007; Puts, Gaulin & Verdolini, 2006). Perception studies have also reported that listeners rate speakers with lower F0 and Fi as being bigger and more dominant than speakers with higher F0 and Fi (Puts et al., 2007; Rendall et al., 2005; Tusing & Dillard, 2000). However, the extent to which F0 and Fi manipulations encode for both dominance and gender characteristics is yet to be systematically explored. The imitation paradigm described in this study could be used to explicitly address this question by asking speakers to express dominance and masculinity both in conjunction and separately (e.g. to sound more dominant, more masculine, dominant and masculine, dominant and feminine). Psychoacoustic studies should also investigate the perceptual relevance of F0 and Fi adjustments in gender and dominance expression and whether the same gestures are perceived differently according to speaker’s and listener’s personality and emotional state, situational context, semantic content and society-specific stereotypes that characterise power and gender relationships.

The present study also explored visible vocal tract length adjustments underlying the observed acoustic manipulations in formant values by providing quantitative measurements of lip movements. We found that, in line with the observed between-sex differences in overall formant spacing, lip spreading and openness were greater in women than in men when speaking normally, suggesting that women speak with a smile. We also found that the majority of participants perceived themselves as spreading their lips more when they feminised their voices than when speaking normally or masculinising them. In line with these self-perceptions, lip measurements revealed that speakers tended to decrease lip spreading from the feminine to the masculine conditions, although significance was only reached when speakers tried to sound as masculine as possible. In contrast, no significant differences across conditions were found for lip openness and ratio. This suggests that lip gestures alone cannot fully account for the observed formant shifts. Indeed, while it was not possible to track vertical laryngeal displacement, more than one third of the participants, and particularly men, reported moving their larynx along the existing sex dimorphism in the experimental conditions and especially when masculinising their voices. It is possible that the enhanced protrusion of the human male larynx, compared to the female larynx, allows male speakers to be more aware of any movement in its position. It is worth noting that the males of several other mammalian species are known to actively lower their larynges during vocalisation in order to extend their vocal tracts and thus exaggerate the vocal expression of their body size (red deer: Reby et al., 2005; fallow deer: Vannoni & McElligott, 2008), pointing at selection pressures underlying the sexual dimorphism of the vocal tract (deer: Fitch & Reby, 2001; humans: Fitch & Giedd, 1999). A recent study also indicates that vocal tract length adjustments affect attributions of physical and social dominance in human males (Puts et al., 2007).

Further investigations should consider more sophisticated techniques to better quantify lip movements (e.g. motion tracking (Yehia, Rubin & Vatikiotis-Bateson, 1998; Kroos, Kuratate & Vatikiotis-Bateson, 2002)), as well as measure laryngeal vertical shifts (e.g. using ultrasound or cine-MRI (Takemoto, Honda, Masaki, Shimada & Fujimoto, 2006)) in order to establish the respective role of such adjustments in the manipulation of vocal tract length to vary the expression of gender or related attributes.

Finally, the observed lip gestures performed to feminise or masculinise the apparent gender of the voice are likely to impact facial expressions and associated gender stereotypes. While Ohala (1984) suggested that the retraction of lip corners to sound smaller and their rounding and protrusion to sound bigger are, respectively, at the origin of the smile and the "o-face" which are common in dominance displays, we propose that individuals feminising their voice are likely to spread their lips, and therefore project a "cheerful", unthreatening face, and those masculinising their voice are likely to round their lips, and therefore project a more "angry", dominant face. Indeed, women tend to smile more than men (Hecht & LaFrance, 1998), possibly following cultural norms (Hall, Carter & Horgan, 2000; LaFrance & Hecht, 1999; LaFrance, Hecht & Paluck, 2003; Stoppard & Gunn Gruchy, 1993).

Future directions

The present study shows that untrained speakers have the spontaneous ability to modify the expression of their gender and related traits through the voice, but does not shed light on their acquisition and use in every day life. We suggest that future studies could (i) extend the imitation paradigm adopted in this study to children and investigate the acquisition and development of sex-typical ways of speaking according to age, (ii) investigate whether children and adults vary the expression of their gender in different settings, and when complying with varying gendered and sex roles within and across different societies, as well as the perceptual relevance of these variations.

Acknowledgments

We are very grateful to Isabelle Charrier, Karen McComb and one anonymous reviewer for their helpful comments on earlier versions of this manuscript.

References

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America, 106*(2), 1054–1063.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America, 69*(1), 231–238.

Boersma, P., & Weenink, D. (2006). PRAAT, a system for doing phonetics by computer(version 5.0.3) (computer program).

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America, 97*(4), 2603–2606.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *The Journal of the Acoustical Society of America, 92*(1), 593–596.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*(5), 997–1004.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*(1), 47–55.

Ericsdotter, C., & Ericsson, A. M. (2009). Gender differences in vowel duration in read Swedish: Preliminary results. *Lund Working Papers in Linguistics, 49*, 34–37.

Fairbanks, G. (1960). *Voice and articulation drillbook*. Harper New York.

Fant G (1960) Acoustic Theory of Speech Production. The Hague: Mouton & Co.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561–568.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. Proceedings of the Royal Society of London. *Series B: Biological Sciences, 26*8(1477), 1669–1675.

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech, 69*(1), 30–57.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology, 6*(4), 219–225.

Goldstein, U. G. (1980). An articulatory model for the vocal tracts of growing children. Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/16118

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26*(4), 351–366.

Günzburger, D., Bresser, A., & Ter Keurs, M. (1987). Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. *Language and Speech, 30*(1), 47–58.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. *Gender and Emotion: Social Psychological Perspectives*, 97–117.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation,/u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America, 123*(5), 2825–2835.

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association, 35*(2), 183–199.

Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin, 24*, 1332–1342.

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication, 9*(4), 299–311.

Henton, C. (1995). Pitch dynamism in female and male speech. *Language & Communication, 15*(1), 43–61.

Hermes, D. J., & Van Gestel, J. C. (1991). The frequency scale of speech intonation. *The Journal of the Acoustical Society of America, 90*(1), 97–102.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099–3111.

Hirano M, Kurita S, Nakashima T (1981) The structure of the vocal folds. In: Stevens K, & Hirano M, editors. *Vocal Fold Physiology*. Tokyo: University of Tokyo Press. pp.33-41.

Kramer, C., Thorne, B., & Henley, N. (1978). Perspectives on language and communication. Signs. Retrieved from http://psycnet.apa.org/psycinfo/1980-25221-001

Kroos, C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Video-based face motion measurement. *Journal of Phonetics, 30*(3), 569–590.

LaFrance, M., & Hecht, M. A. (1999). Option or obligation to smile: The effects of power and gender on facial expression. Retrieved from http://psycnet.apa.org/psycinfo/1999-04108-002

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. Psychological Bulletin, 129(2), 305.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*, 1455.

Linke, C. E. (1973). A study of pitch characteristics of female voices and their relationship to vocal effectiveness. *Folia Phoniatrica et Logopaedica, 25*(3), 173–185.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica, 25*(2), 119–125.

McConnell-Ginet, S. (1978). Intonation in a man's world. Signs, 3(3), 541–559.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech, 50*(1), 125–142.

Munson, B., & Babel, M. (2007). Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass, 1*(5), 416–449.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*(1), 1–16.

Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. *Multilingualism, Second Language Learning, and Gender*, 231–254.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America, 109*, 2988.

Peterson, G. E., & Barney, H. L. (2005). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 175–184.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America, 116*, 1905.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*(4), 2201–2212.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. Evolution and Human Behavior, 26(5), 388–397.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*(4), 283–296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340–344.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*(3), 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences, 272*(1566), 941–947.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America, 117*(2), 944–955.

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The Queen's English: an alternative, biosocial hypothesis for the distinctive features of 'gay speech'. *Archives of Sexual Behavior, 37*(1), 188–204.

Rietveld, A. C. M., & Gussenhovent, C. (1985). On the relation between pitch excursion size and prommence. *Journal of Phonetics, 13*, 299–308.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects,* 74–84.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and agea). *The Journal of the*

Acoustical Society of America, 118(5), 3177–3186.

Simpson, A. P. (2003). Possible articulatory reasons for sex-specific differences in vowel duration. *In Proc. 6th International Seminar on Speech Production,* Sydney.

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, US: MIT Press.

Stoppard, J. M., & Gruchy, C. D. G. (1993). Gender, context, and expression of positive emotion. Personality and Social Psychology Bulletin, 19(2), 143–150.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., & Fujimoto, I. (2006). Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *The Journal of the Acoustical Society of America*, *119*(2), 1037–1049.

Thorsen, N. (1980). Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *14*, 1–29.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. *Human Communication Research*, 26(1), 148–171.

Vannoni, E., & McElligott, A. G. (2008). Low Frequency Groans Indicate Larger and More Dominant Fallow Deer (Dama dama) Males. *PLoS ONE*, 3(9), e3113.

Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language, and Hearing Research : JSLHR, 50*(6), 1510–1545.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging studya). *The Journal of the Acoustical Society of America*, *125*(3), 1666–1678.

Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association, 26*(1), 23–40.

Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication, 32*(4), 267–285.

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders, 55*(1), 43–50.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*(1), 23–43.

Appendix A

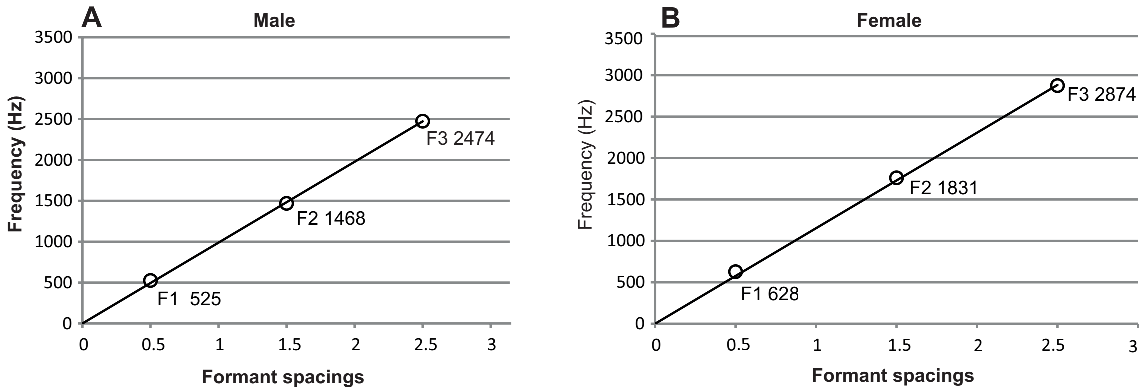

Figure 4.5.3. Illustration of the fitness of the method used to estimate overall formant spacing. Frequency values of F1, F2 and F3 for male (A) and female (B) adult (>19 years old) speakers as measured in Lee and colleagues (1999) plotted against (2i−1)/2 increments of the formant spacing as predicted by a uniform vocal tract model. Formant spacing (ΔF) can be estimated as the slope of the linear regression of observed Fi over the expected formant positions (with intercept set to 0). The apparent Vocal Tract Length (aVTL expressed in centimetres) can be calculated as aVTL = c/2ΔF. The values of *ΔF* reported in the figures correspond to aVTL values of 17.71cm for male speakers and 14.95cm for female speakers, which are comparable to anatomical vocal tract lengths in adult men and women (men: 18cm, women: 15cm (Vorperian et al., 2008)). This illustrates that, while *ΔF* estimated in this way is sensitive to vowel-specific variation in vocal tract configuration, at supra-segmental level it provides an estimate of the overall linear scaling of the formants, which is a reliable estimate of the average vocal tract length of the speaker.

Appendix B

Table 4.5.2

ANOVA table for the acoustic parameters in vowel task (N=31)

Acoustic parameters	Condition		Sex		Sex x Condition	
	F	*p*	*F*	*p*	*F*	*p*
F0*mean*	55.05	< .001*	118.75	< .001*	1.61	.215
F0*CV*	1.17	.318	0.14	.713	1.30	.280
F1	10.30	< .001*	50.58	< .001*	5.40	.011*
F2	25.76	< .001*	67.50	< .001*	2.96	.060
F3	18.58	< .001*	39.98	< .001*	1.03	.349
F4	29.27	< .001*	60.09	< .001*	4.78	.024*
ΔF	30.33	< .001*	73.13	< .001*	2.48	.114

Note. F-ratio (F) and p-value (*p*) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (ΔF). Significant effects are indicated with an asterisk.

Table 4.5.3

ANOVA table for the acoustic parameters in sentence task (N=32)

Acoustic parameters	Condition		Sex		Sex x Condition	
	F	*p*	*F*	*p*	*F*	*p*
F0*mean*	54.16	< .001*	139.32	< .001*	0.97	.351
F0*CV*	3.61	.044*	17.15	< .001*	1.47	.240
F1	4.73	.018*	14.39	.001*	6.71	.005*
F2	14.09	< .001*	23.92	< .001*	1.73	.196
F3	13.91	< .001*	27.20	< .001*	2.18	.142
F4	47.71	< .001*	72.39	< .001*	6.15	.011*
ΔF	41.76	< .001*	62.28	< .001*	2.01	.162

Note. F-ratio (F) and p-value (*p*) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (ΔF). Significant effects are indicated with an asterisk.

Table 4.5.4

ANOVA table for the acoustic parameters in passage task (N=32)

Acoustic parameters	Condition		Sex		Sex x Condition	
	F	*p*	*F*	*p*	*F*	*p*
F0*mean*	38.26	< .001*	186.65	< .001*	0.69	.506
F0*CV*	4.68	.018*	4.93	.034*	2.16	.134
F1	13.58	< .001*	17.83	< .001*	4.15	.030*
F2	17.18	< .001*	52.56	< .001*	1.51	.231
F3	21.71	< .001*	43.09	< .001*	1.67	.204
F4	22.73	< .001*	88.61	< .001*	0.52	.561
ΔF	23.35	< .001*	81.49	< .001*	0.97	.365

Note. F-ratio (F) and p-value (*p*) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (ΔF). Significant effects are indicated with an asterisk.

Table 4.5.5

Mean and Standard Deviation (SD) of female speakers' acoustic parameters

Acoustic parameters	Condition					
	Masc		Neutral		Fem	
All vowels (N=14)	mean	SD	mean	SD	mean	SD
F0*mean*	185.6	25.3	202.41	22.9	256.6	55.4
F0*CV*	.11	.05	.10	.06	.13	.08
F1	568.5	59.3	648.0	92.0	667.2	90.9
F2	1795.8	128.8	1924.6	101.4	1948.5	109.4
F3	2795.6	166.9	2917.0	155.0	2964.7	121.8
F4	3938.6	210.3	4090.1	192.7	4123.7	153
ΔF	1131.1	58.9	1181.9	50.1	1195.2	43.7
Sentence (N=15)						
F0*mean*	178.8	22.4	196.2	30.2	256.7	47
F0*CV*	.19	.10	.25	.10	.21	.08
F1	486.4	75.3	512.1	69.4	592.3	72.0
F2	1827.5	102.8	1926.4	136.2	2029.1	183.6
F3	2642.6	240.6	2810.6	174.3	2899.9	203.4
F4	3847.5	243.2	4021.7	209.6	4132.8	202.2
ΔF	1098	72.2	1154.7	56.3	1193.1	55.8
Passage (N=15)						
F0*mean*	184.6	25.7	188.9	25.2	238.5	42.6
F0*CV*	.18	.10	.23	.10	.23	.10
F1	584.7	48.4	634.9	52.7	646	63.8
F2	1761.4	82.2	1831.7	93.6	1851	104.3
F3	2870.1	128.2	2983.9	134.2	3020.3	158.4
F4	3967.8	125.9	4075.2	137.1	4133.6	187.6
ΔF	1142.7	40.5	1180.4	44.5	1196.1	56.9

Note. Mean and SD values (Hz) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (*ΔF*). "Masc" and "Fem" represent the masculinised and feminised conditions.

Table 4.5.6

Mean and Standard Deviation (SD) of male speakers' acoustic parameters

Acoustic parameters	Condition					
	Masc		Neutral		Fem	
All vowels (N=17)	mean	SD	mean	SD	mean	SD
F0*mean*	103.2	11.9	107.6	13.78	152.3	37.4
F0*CV*	.11	.04	.11	.05	.11	.07
F1	474.8	65.7	472.7	45.5	499.4	71.8
F2	1579.2	110.4	1619.9	88.4	1682.4	96.8
F3	2559.0	138.2	2609.1	126.5	2717.9	153.8
F4	3369.6	239.8	3508.9	236.8	3743.9	237.5
ΔF	990.3	58	1022.4	55	1079.6	59.9
Sentence (N=17)						
F0*mean*	103.8	13.1	110.6	11.3	162.2	47.7
F0*CV*	.10	.06	.15	.10	.20	.05
F1	460.5	168.1	396.5	49.3	430.4	93.3
F2	1660.9	164.7	1697.9	155.4	1758.2	183,8
F3	2424.8	199	2436	158.8	2572.5	254.9
F4	3199.4	160.4	3357.2	185,3	3731.9	349.2
ΔF	951.5	55.5	980.3	52.9	1064.1	89.7
Passage (N=17)						
F0*mean*	105.4	11.2	106	10.3	145.6	39.1
F0*CV*	.16	.06	.16	.04	.20	.04
F1	523.1	73.9	527.6	70.1	548.1	75.3
F2	1583.9	78.3	1606.5	65.3	1660.8	109.3
F3	2662.7	84.4	2701.3	64.9	2788.8	152
F4	3591.0	101.6	3662.2	112	3770.9	173.2
ΔF	1041.1	28.3	1059.3	21.6	1092.1	52.5

Note. Mean and SD values (Hz) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (*ΔF*). "Masc" and "Fem" represent the masculinised and feminised conditions.

Table 4.5.7

Within-sex contrasts for the acoustic parameters across conditions in female speakers

Acoustic parameters	Contrasts			
	Neutral vs. Masc		Neutral vs. Fem	
All vowels (N=14)	F	P	F	p
F0*mean*	14.31	.002*	24.80	<.001*
F0*CV*	0.26	.619	5.33	.038*
F1	10.17	.007*	0.34	.569
F2	17.10	.001*	1.59	.229
F3	10.56	.006*	2.57	.133
F4	20.60	.001*	0.99	.002*
ΔF	26.17	< .001*	2.15	.166
Sentence (N=15)				
F0*mean*	5.99	.028*	18.26	.001*
F0*CV*	2.49	1.370	1.622	.224
F1	3.21	.095	24.89	< .001*
F2	15.83	.001*	11.98	.004*
F3	13.45	.003*	4.58	.050
F4	19.72	.001*	6.81	.021*
ΔF	32.32	< .001*	12.32	.003*
Passage (N=15)				
F0*mean*	.86	.370	24.92	< .001*
F0*CV*	6.81	.021*	.040	.840
F1	20.23	.001*	.790	.388
F2	13.32	.003*	.690	.420
F3	20.96	< .001*	1.49	.242
F4	11.02	.005*	2.08	.172
ΔF	15.81	.001*	1.78	.210

Note. F-ratio (F) and p-value (*p*) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1−F4) and formant spacing (*ΔF*). Significant effects are indicated with an asterisk. “Masc” and “Fem” represent the masculinised and feminised conditions.

Table 4.5.8

Within-sex contrasts for the acoustic parameters across conditions in male speakers

Acoustic parameters	Contrasts			
	Neutral vs. Masc		Neutral vs. Fem	
All vowels (N=17)	F	p	F	p
F0*mean*	5.38	.034*	36.95	< .001*
F0*CV*	.01	.919	0.01	.942
F1	.07	.798	4.18	.058
F2	5.75	.029*	7.08	.017*
F3	7.45	.015*	6.71	.020*
F4	26.17	< .001*	12.17	.003*
ΔF	22.69	< .001*	10.96	.004*
Sentence (N=17)				
F0*mean*	8.51	.010*	24.33	< .001*
F0*CV*	1.83	.195	1.28	.275
F1	2.22	.155	1.45	.246
F2	.86	.367	3.76	.070
F3	.17	.688	5.93	.027*
F4	20.9	< .001*	28.3	< .001*
ΔF	7.93	.012*	23.38	< .001*
Passage (N=17)				
F0*mean*	.84	.776	14.48	.002*
F0*CV*	3.12	.096	4.11	.060
F1	.40	.537	3.98	.064
F2	6.43	.022*	7.52	.014*
F3	7.64	.014*	7.46	.015*
F4	13.46	.002*	8.58	.010*
ΔF	13.77	.002*	8.60	.010*

Note. F-ratio (F) and p-value (*p*) for: mean fundamental frequency (F0*mean*), coefficient of variation (F0*CV*), first four formant frequencies (F1-F4) and formant spacing (*ΔF*). Significant effects are indicated with an asterisk. "Masc" and "Fem" represent the masculinised and feminised conditions.

Table 4.5.9

ANOVA table for the vowel formant frequencies

All vowels	Condition F	p	Vowel F	p	Condition x Vowel F	p
Women (N=14)						
F1	12.48	< .001*	59.14	< .001*	.50	.950
F2	11.53	< .001*	72.53	< .001*	.53	.930
F3	11.99	< .001*	12.49	< .001*	.48	.960
F4	12.46	< .001*	2.41	.016*	.68	.811
Men (N=17)						
F1	3.53	.030*	87.71	< .001*	1.06	.394
F2	8.26	< .001*	178.21	< .001*	.65	.841
F3	16.92	< .001*	27.94	< .001*	.56	.918
F4	50.27	< .001*	7.36	< .001*	.45	.969

Note. Significant effects are indicated with an asterisk.

Table 4.5.10

Mean, standard deviation (SD) and contrasts for Lip Spreading (LS), Lip Openness (LO) and Lip Ratio.

All vowels	Condition Masc		Neutral		Fem		Contrasts Neutral vs. Masc		Neutral vs. Fem	
Women (N=14)	mean	SD	mean	SD	mean	SD	F	p	F	p
LS	86.5	10.7	88.4	9.6	90.7	9	5.71	.044*	4.11	.077
LO	18.7	2.0	21.1	1.5	20.5	2.2	3.94	.082	0.29	.603
Lip ratio	5.4	2	4.7	1.3	5.2	1.7	3.5	.098	2.34	1.650
Men (N=17)										
	mean	SD	mean	SD	mean	SD	F	p	F	p
LS	66.7	17.6	69.3	19.8	69.4	18.7	6.5	.024*	.07	.791
LO	14.6	1.2	14.5	.9	15	1.3	.002	.968	.32	.581
Lip ratio	5.7	1.5	5.4	1.1	5.4	1.3	.78	.392	.10	.758

Note. Significant effects are indicated with an asterisk. “Masc” and “Fem” represent the masculinised and feminised conditions.

Study 6: Confirming the Perceptual Relevance of Voice Gender Control by Children and Adult Speakers

Abstract

Our voices are sexually dimorphic: men speak with markedly lower fundamental frequency (F0) and lower, more closely spaced vocal tract resonances (ΔF) than women, and pre-pubertal boys speak with lower ΔF than girls. While this acoustic dimorphism has a strong biological basis, recent research has shown that from childhood speakers can also manipulate F0 and ΔF to accentuate or de-emphasise their perceived gender (by lowering F0 and ΔF to masculinise their voices, and raising F0 and ΔF to feminise them). However, the perceptual relevance of these behavioural adjustments remains to be investigated. Here, we asked adult listeners to characterise the gender of pre-pubertal and adult speakers as they spoke normally, as well as when trying to sound as masculine or as feminine as possible. Results revealed that adults consistently rated lower-pitched, more resonant voices as belonging to more masculine speakers than higher-pitched, less resonant voices. These results confirm that voice gestures performed by speakers in order to vary the expression of their voice gender are perceptually relevant, providing further support for the role of vocal behaviours in voice gender expression.

Introduction

We can reliably identify the gender of speakers from listening to their voices only. This ability is present from a very early age: at seven months, infants are able to consistently distinguish between pre-pubertal boys and girls (Bahrick, Netto, & Hernandez-Keif, 1998), and between adult male and female voices (Miller, Lurye, Zosuls, & Ruble, 1982). By the age of six, child listeners' performance approaches the high levels of accuracy shown by adult listeners (Hillenbrand & Clark, 2009).

In adult voices, gender identification is mainly signalled by two key sexually dimorphic traits, voice fundamental frequency (F0) and the global pattern of vocal tract resonances, or formant spacing (ΔF), which are both lower in men's voices than in women's (Titze, 1994). These voice differences are in turn largely dependent on size differences in the voice production mechanisms: F0 and ΔF are, respectively, inversely related to the length of vocal folds and vocal tract, which are on average longer in men than in women. However, emerging evidence from cross-linguistic studies (Johnson et al., 2006; Ohara et al., 1999) and acoustic investigations of homosexual and heterosexual voices (Munson, 2007; Rendall, Vasey, & McKenzie, 2008) suggests that gender differences in the voice frequency parameters of adults cannot solely be accounted for by biological factors. Even more intriguing is the observed dimorphism in the pre-pubertal voice: acoustic data consistently report that pre-pubertal boys speak with lower ΔF than girls (Bennett, 1981; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde & Ashmead, 2001), and correspondingly, adult listeners tend to use ΔF when assigning children's gender (Perry, et al., 2001; Cartei & Reby, 2013), despite the absence of significant sex differences in the overall length of children's vocal tracts prior to the onset of puberty. Taken together, these results indicate that biological differences between males and females are not sufficient to explain variation in voice gender expression, raising the possibility that some of this variation may be due to vocal behaviour. Initial support for this hypothesis was reported in a study showing that adult speakers were able to spontaneously lower their F0 and ΔF when asked to sound more masculine and to raise these components when asked to sound more feminine (Cartei Cowles, & Reby, 2012). Similar abilities were subsequently observed in pre-pubertal children: when asked to sound as much as possible "like a boy", six to nine year olds would lower their F0 (girls only) and ΔF (both genders), while raising their F0 (boys only) and ΔF (both genders) to sound as much as possible "like a girl" (Cartei, Cowles, Banerjee & Reby, 2013). Both studies suggest that from childhood speakers may use a "gender code", making a conventionalised use of the existing sex dimorphism in the frequency components of their voice (ΔF in children and adults, F0 in adults) to vary the expression of gender and related (e.g. masculinity, femininity) characteristics. While these studies show that speakers control their voices in a way that accentuates or downplay their gender attributes, thus providing evidence for the "gender code" at the production level, the "gender code" also implies that, at a perceptual level, listeners should be attentive to such voice adjustments and be affected by them when characterising speakers' gender attributes. Indeed, psychoacoustic studies have shown that listeners are sensitive to artificial manipulations of F0 and ΔF, with lower-pitched (in adults) and more resonant voices (in both children and adults) being consistently rated as more masculine than their higher-pitched, less resonant versions (Pisanski, Mishra & Rendall, 2012; Pisanski & Rendall, 2011; Cartei & Reby, 2013). However, no study has so far investigated whether speakers' own behavioural F0 and ΔF adjustments to vary their perceived gender and related attributes have a perceptual relevance. This is therefore the main focus of the present study. In Experiment 1 we asked listeners to rate pre-pubertal children's voices on a gender scale (masculine boy to feminine girl) as they spoke in their normal speaking voice (neutral condition), and when trying to sound like a boy (masculinised version) or girl (feminised version) as much as possible. In Experiment 2, we asked listeners to rate the masculinity of adult speakers (from "very masculine" to "not at all masculine") as they spoke in their normal speaking voices (neutral condition), and when sounding as masculine (masculinised condition) and as feminine (feminised condition) as possible. We predicted that masculinised voices of child and adult speakers would be perceived as more masculine than neutral or feminised voices. Similarly, we predicted that feminised voices would be perceived as more feminine than neutral or masculinised voices.

Experiment 1. Perception of voice gender manipulations in pre-pubertal children's voices

Methods

Voice stimuli. The voice recordings of 15 boys and 19 girls (Mean age = 7.04, *SD* = 11.1) were taken from a previous production experiment (Cartei et al., 2013), where children read words out loud (/AE/ "hat", /EH/ "bed", /ER/ "bird", /IY/ "feet", /IH/ "pig", /AH/ "duck", /AA/ "box", /UH/ "book", /UY/ "boot"), first in their normal speaking voice (neutral condition) and then sounding as much as possible like a boy (masculinised condition) or a girl (feminised condition). Mean and standard deviations for F0 and ΔF of child speakers in each condition (Table 4.6.1) were estimated using a custom PRAAT script (Cartei et al., 2013 for details). All vowels from the list of words uttered by the same speaker were concatenated in each voice condition with 50 ms silent interval in between, and standardised to 65 dB, resulting in 102 stimuli in total (three stimuli (masculinised, neutral, feminised) x 34 speakers).

Table 4.6.1

Mean (SD) in Hz for fundamental frequency (F0) and format spacing (ΔF) of boys and girls in the masculinised, neutral and feminised conditions

Speakers' Sex	Acoustic parameters	Masculinised	Neutral	Feminised
Boys (N=19)	F0	243.5 (32)	249.6 (29)	307.2 (62)
	ΔF	1284 (69)	1313 (68)	1355 (80)
Girls (N=15)	F0	234.6 (30)	249.1 (26)	270.2 (50)
	ΔF	1301 (67)	1355 (46)	1389 (53)

Note. Values as published in Cartei and colleagues (2013). Boys spoke with a significantly lower ΔF (3.2%) than girls, but same F0. Compared to their normal speaking ("neutral") voices, boys' mean F0 and ΔF were 2.4% and 2.2% lower in the masculine condition (albeit only the lowering of ΔF was significant), and 23.2% and 3.2% higher in the feminine condition. Compared to their neutral voices, girls' mean F0 and ΔF were respectively, 5.8% and 3.9% lower in the masculine condition, while their F0 and ΔF were, respectively, 8.5% and 2.5% higher in the feminine condition (albeit only the upward shift in ΔF was significant). All percentage changes were calculated from the means using the formula: $((V_2 - V_1) / |V_1|) * 100$).

Participants. 245 second-year Psychology students (94 males and 151 females, mean age = 21, *SD* = 2.1) from Sussex University undertook the experiment as part of their Cognitive Psychology level two module. No participants reported a history of hearing impairments. Written informed consent forms were obtained from all participants.

Experimental Procedure. Participants completed the experiment in a sound-controlled room at the University of Sussex. Before the experiment began, each participant was sat in front of a computer screen and instructed to wear *Dynamode dh-660mv* headphones set at a default volume level (65%), adjusting their volume if needed. Participants then loaded the experiment on their computer using a custom script written in PRAAT v.5.20 (Boersma & Weenink, 2011). Participants followed the on-screen instructions, informing them that they would hear through their headphones a series of vowels spoken by different child speakers, and would be asked to rate the speaker's voice on a scale from 1 to 7 ("Rate the voice of the speaker on a scale from 1 to 7") after listening to each set. For each voice stimulus, the scale was represented on the screen by a set of seven buttons, labelled as 1 = masculine boy, 2 = boy, 3 = feminine boy, 4 = neutral, 5 = masculine girl, 6 = girl, 7 = feminine girl. Once the rating was made, the script would automatically present the next stimulus. The voices were presented once, with their order being randomised. Each participant thus completed 102 trials (three voice stimuli x 34 different speakers), with scheduled rest-break intervals after 35 voice stimuli.

Statistical Analysis. Reliability between raters on perceived gender was estimated from raw scores using Cronbach's Alpha (α (Cronbach, 1951)). This method involves measuring the correlation between each individual listener's rating for each stimulus with the group mean of all the other listeners. A α value of .80 or above indicates that listeners agree very well with one another (Bohrnstedt, 1970 as cited in Feinberg, Jones, Little, Burt, & Perrett, 2005). Mean ratings were calculated by averaging gender ratings for each speaker in each condition (masculinised, feminised, neutral). A Linear Mixed Model with voice condition (masculinised, neutral, feminised) and speakers' sex (male, female) as fixed factors and the speaker's identifier (speaker ID) as random factor, was run to assess whether those factors and their interactions had a significant effect on listeners' mean ratings. Pairwise comparisons were performed to compare individual levels of the voice condition factor. All analyses used two-tailed probability estimates and were run using SPSS v.20.

Results

Inter-related reliability was high ($\alpha = .85$), indicating that listeners agreed well on ratings. Mean ratings by speaker's sex and condition are illustrated in Figure 4.6.1. The main effect of sex was not significant, $F(1, 32) = 2.55$, $p = .121$. Voice condition had a significant main effect on ratings of boys' and girls' voices, $F(2, 64) = 50.8$, $p < .001$. Pairwise comparisons revealed that mean rating scores in each condition were significantly different from the other two conditions, *ps* <.001: masculinised voices of both sexes received the lowest gender ratings (boys: $M = 3.2$, $SE =.24$, girls: $M = 3.2$, $SE = .22$, close to the "feminine boy" score of the scale), while feminised voices received the highest ratings (boys: $M = 4.4$, $SE = .24$; girls: $M = 4.8$, $SE = .22$, close to the "masculine girl" score of the scale). The effect of the interaction between condition and sex was significant $F(2, 64) = 6.4$, $p = .003$: in the neutral condition, boys' voices were rated as significantly more masculine ($M = 3.5$, $SE = .24$, between the "feminine boy" and "neutral" scores) than girls ($M = 4.5$, $SE = .22$, between the "neutral" and "masculine girl" scores), while they received similar ratings to girls in the other two conditions (Figure 4.6.1).

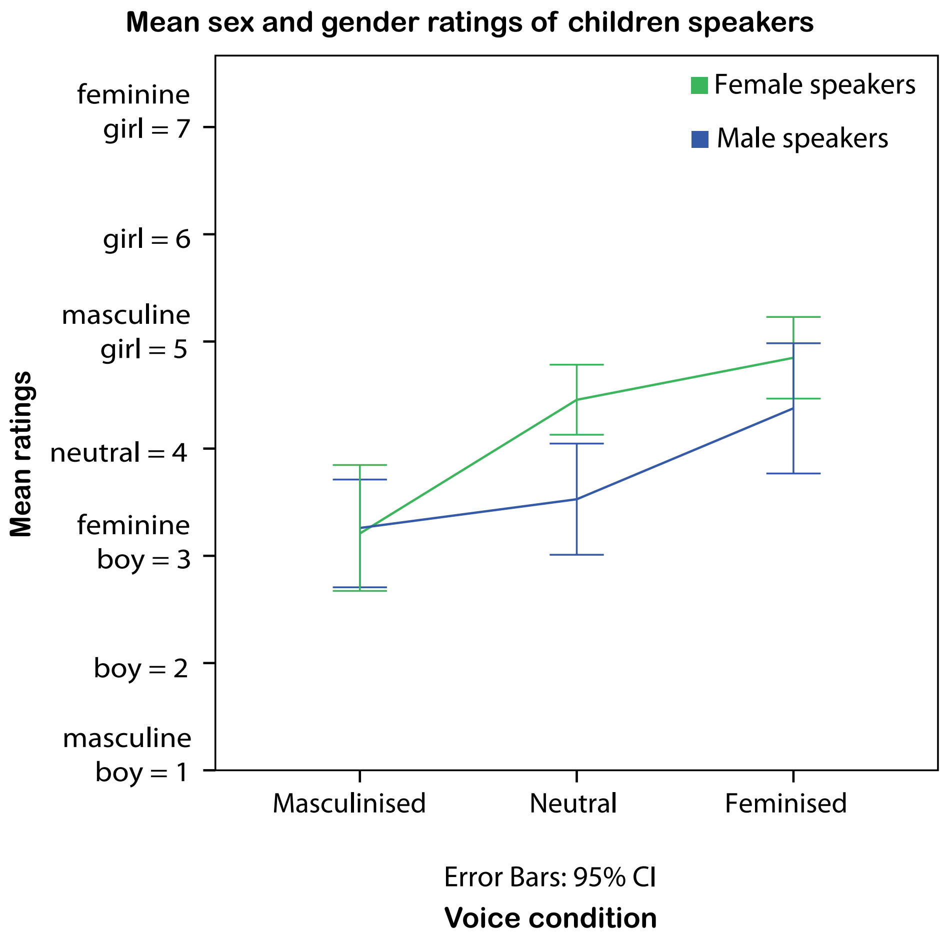

Figure 4.6.1. Mean sex and gender ratings of masculinised, neutral and feminised voices in pre-pubertal speakers. When speaking normally, boys' voices received significantly lower ratings than girls' voices, Ratings in the other two conditions were not significantly different between the sexes: ratings of boys' and girls' masculinised voices received the lowest ratings (close to the "feminine boy" – 3 score), while their feminised voices received the highest ratings (close to "masculine girl" – 5 score).

Experiment 2. Perception of voice gender manipulations in adults' voices

Methods

Voice stimuli. Ten male and ten female speakers were randomly selected from the speakers' database in Cartei and colleagues (2012), where speakers were asked to read CVC words (/AE/ "had", /EH/ "head", /ER/ "heard", /IY/ "heed", /IH/ "hid", /AH/ "hud", /AA/ "hod", /UH/ "hood", /UY/ "who'd") out loud in their normal speaking voice (neutral condition) and then trying to sound as masculine (masculinised condition) or as feminine (feminised condition) as possible. Means and standard deviations for F0 and ΔF for this subset of speakers in each condition were estimated using a custom PRAAT script (Cartei et al., 2012 for details). Values (Table 4.6.2) are in line with those reported in the bigger sample (Cartei et al., 2012). All vowels from the list of single-syllable words uttered by the same speaker were concatenated in each voice condition with 50-ms silent interval in between, and standardised to 65 dB, resulting in 60 stimuli in total (three stimuli (masculinised, neutral, feminised) x 20 participants).

Table 4.6.2

Mean (SD) in Hz for fundamental frequency (F0) and format spacing (ΔF) of men and women in the masculinised, neutral and feminised conditions

Speakers' sex	Acoustic parameters	Masculinised	Neutral	Feminised
Men (N=10)	F0	103.5 (12.5)	110.2 (14.0)	154.7 (38.7)
	ΔF	986.5 (54.3)	1024.3 (60.1)	1078.9 (65.0)
Women (N=10)	F0	186.3 (25.8)	207.5 (21.2)	252.3 (43.9)
	ΔF	1133.6(60.2)	1185.5 (49.9)	1202.3 (37.0)

Note. Compared to their normal voices, female speakers lowered their F0 by 11.4% and ΔF by 4.6% when masculinising their voices, and raised F0 by 21.5% and ΔF by 1.41% when feminising them. Male speakers lowered F0 by 6.4% and ΔF by 3.8% when masculinising their voices, and raised F0 by 40.4% and ΔF by 5.3% when feminising them. These values are in line with the adjustments reported in Cartei and colleagues (2012) across speakers, whereby females significantly lowered their F0 by 9.1% and ΔF by 4.5% when masculinising their voices, and significantly raised F0 by 26.8%, and ΔF by 1.1% when feminising them, while males significantly lowered F0 by 4.3% and ΔF by 3.2% when masculinising their voices, and significantly raised F0 by 41.5% and raised ΔF by 5.6% (albeit not significantly) when feminising them. All percentage changes were calculated from the means using the formula: $((V_2 - V_1) / |V_1|) * 100$).

Participants. Sixty-three second-year Psychology students (21 males and 42 females, mean age = 21, *SD* = 2.4) from Sussex University undertook the experiment as part of their Cognitive Psychology level two module. No participants reported a history of hearing impairments. Written informed consent forms were obtained from all participants.

Experimental Procedure. Participants were collectively tested in a sound-attenuated room. Each participant sat in front of a computer screen wearing *Dynamode dh-660mv* headsets set at a default volume (65%). Voice stimuli and instructions were presented via a custom PRAAT script, following the procedure in Experiment 1. Male and female voice stimuli were presented together. Participants rated each voice for masculinity (e.g. "How masculine does the speaker sound?"), using a seven-point scale represented on-screen by a set of seven buttons (from 1 = very masculine to 7 = not at all masculine). Each participant thus completed 60 trials (three voice stimuli x 20 different speakers), with one scheduled rest-break interval after 30 voice stimuli.

Statistical Analysis. Reliability between raters on perceived masculinity was estimated from raw scores using Cronbach's Alpha (α) separately for male and female speakers. Mean ratings were calculated by averaging masculinity ratings for each speaker in each condition (masculinised, feminised, neutral). A Linear Mixed Model with voice condition (masculinised, neutral, feminised) and speakers' sex (male, female) as fixed factors and speaker's identifier (speaker ID) as random factor, was run to assess whether those factors and their interactions had a significant effect on listeners' mean ratings. Pairwise comparisons were performed to compare individual levels of the voice condition factor. All analyses used two-tailed probability estimates and were run using SPSS v.20.

Results

Inter-rater reliability was high, with $\alpha = 0.84$ for masculinity ratings of female speakers and $\alpha = .81$ for masculinity ratings of male speakers, indicating a good level of agreement between raters. Mean ratings by speaker's sex and condition are illustrated in Figure 4.6.2. There was a significant main effect of speakers' sex, $F(1, 18) = 314.07$, $p < .001$: overall, male voices were rated as significantly more masculine than female voices. There was also a significant main effect of condition on listeners' ratings of masculinity, $F(1, 18) = 314.064$, $p < .001$. Pairwise comparisons revealed that ratings in each condition were significantly different from the other two, *ps* < .001. In men, masculinised voices received the lowest ratings ($M = 1.6$, $SE = .07$), closest to the extreme masculine end of the scale (1="very masculine" score), their feminised voices received the highest ratings (*M* = 4.3, *SE* = .07), while ratings of their normal speaking voices ("neutral condition") were below the middle point of the scale (*M* = 2.2, *SE* = .07). Similarly, women's masculinised voices received the lowest ratings (*M* = 2.5, *SE* = .07), their feminised voices received the highest ratings (*M* = 4.8, *SE* = .07) and their normal speaking voices received middle ratings (*M* = 3.9, *SE* = .07). The interaction effect between speakers' sex and condition was also significant, $F(2, 36) = 45.32$, $p < .001$: the difference in masculinity ratings between the men's and women's voices was greatest for normal speaking voices (1.7 mean difference score), followed by masculinised (0.9 mean difference score) and feminised (.05 mean difference score) voices.

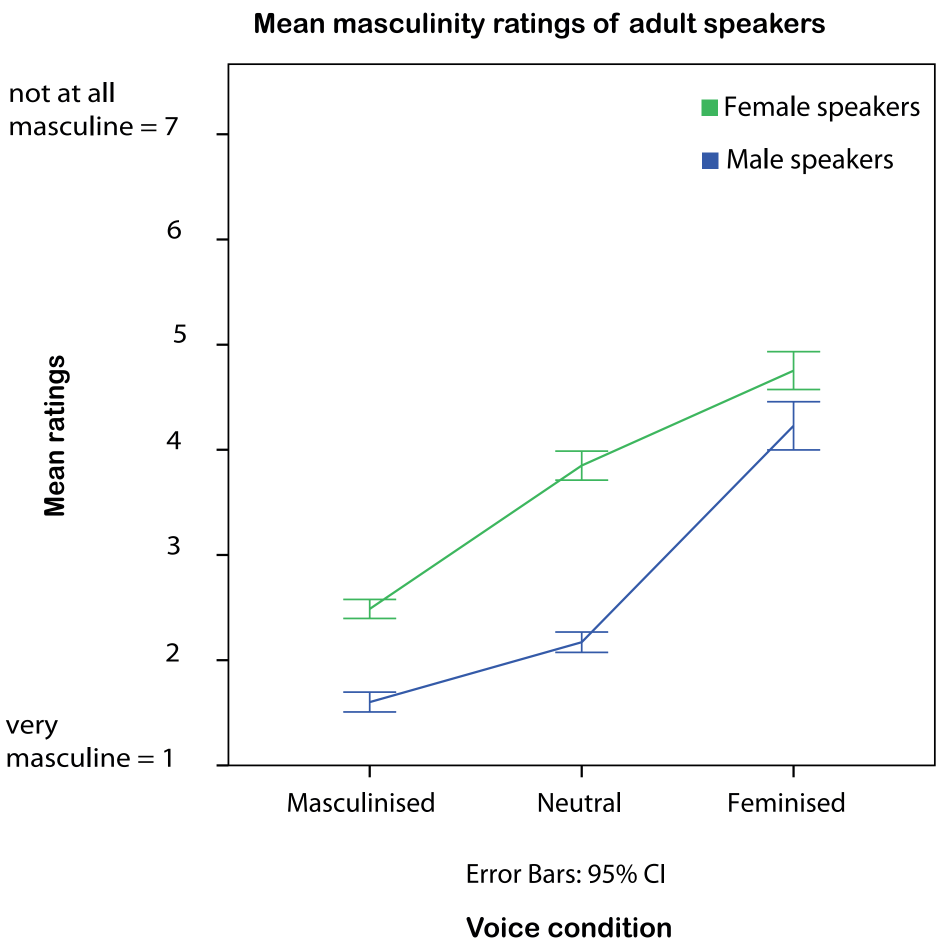

Figure 4.6.2. Mean masculinity ratings in adult speakers. Men's voices were rated as more masculine than women's voices across conditions. This difference was at its highest for normal speaking voices (mean ratings in men: 4.3 and women: 4.8). Masculinised voices from both sexes were rated as more masculine (mean ratings in men: 1.6 and women: 2.5) than the other two, while feminised voices from both sexes were rated as less masculine than the other two (mean ratings in men: 4.3 and women: 4.8).

Discussion

In line with our initial hypothesis, the present results showed that pre-pubertal child speakers (Experiment 1) were perceived to be more masculine when masculinising their voices, and more feminine when feminising their voices, than when speaking normally. Similar results have been reported in the first and only psychoacoustic study of gender perception in pre-pubertal voices (Cartei & Reby, 2013), where linear increments of ΔF within the natural range of children's voices were found to proportionally affect listeners' attributions of speakers' gender (from masculinity to femininity). Indeed, the magnitude of the present ΔF shifts (2%−4%) relative to children's normal speaking voices is comparable to the 2% ΔF manipulations used in Cartei and Reby (2013), suggesting that listeners are finely attuned to subtle ΔF variation in pre-pubertal voices when making gender-related assessments.

We also found that, when speaking normally, boys received lower ratings than girls, providing further evidence that children's voices cue the speaker's gender, in line with previous perceptual (Perry et al., 2001) and acoustic (Cartei et al., 2013; Whiteside, 2001) investigations. According to the observed ΔF values, boys' manipulated voices should have also received lower gender ratings than girls': ΔF values for boys' masculinised (1284Hz) and feminised (1355Hz) voices fall below the perceived sex boundary reported by Cartei and Reby for boys (ΔF of 1396Hz), while girls' ΔF values for their masculinised (1301Hz) and feminised (1389Hz) voices are above the perceived sex boundary for girls (ΔF of 1294Hz (Cartei & Reby, 2013)). Yet, the frequency shifts spontaneously performed by child speakers to sound more "like a boy" or "like a girl", produced a child voice that was systematically perceived as belonging to the opposite sex: in both sexes, ratings of masculinised voices were close to the "feminine boy", while ratings of feminised voices were close to the "masculine girl" score. It is worth noting that the voice stimuli used in Experiment 1 were derived from the same speaker database used in Cartei and colleagues (2013), who found that, in addition to ΔF, children significantly shifted their F0 when imitating the opposite sex. These observations raise the interesting possibility that F0 may also affect gender identification from pre-pubertal voices, despite the lack of dimorphism in this parameter before puberty (Titze, 1994).

More generally, in the absence of overall differences in children's vocal anatomy before puberty (Vorperian et al., 2009; Vorperian et al., 2011), these results lend further support to the hypothesis that children's acoustic dimorphism may have a behavioural, rather than anatomical, origin (Cartei et al., 2013; Lee, et al., 1999; Sachs, Lieberman & Erickson, 1973).

Listeners' assessments of voice masculinity in adult speakers (Experiment 2) showed that men's voices were consistently perceived to be more masculine than women's, as expected from the fact that voice F0 and ΔF values in both men and women remained within the range of their own sex across conditions. Moreover, in line with our hypothesis, adult voices of both men and women were perceived to be more masculine when masculinising their voices, and less masculine when feminising their voices, than when speaking normally. Similarly, previous psychoacoustic studies have shown that adult voices with artificially lowered F0, ΔF or both, are consistently rated as more masculine than those with such parameters raised (Assman, Dembling & Nearey, 2006; Feinberg et al., 2006; Feinberg et al., 2008; Pisanski et al., 2012; Pisanski & Rendall, 2011). However, the relative role of F0 and ΔF in listeners' gendered attributions of adult voices remains to be directly examined. Here we note that F0 and ΔF shifts performed by speakers to masculinise or feminise their voices were above frequency discrimination thresholds (1.5%–9%, see Kewley-Port, Li, Zheng, & Neel, 1996 for a review), although the magnitudes of ΔF shifts were far smaller than those of the F0, and mostly below the minimum difference reported to affect listeners' perceived masculinity (4%–12% (Pisanski & Rendall, 2011)). Future psychoacoustic studies could assess the extent to which speakers' behavioural adjustments of either voice feature independently affects perceptual ratings of gender-related dimensions, for example by independently resynthesising voice F0 and ΔF according to the magnitude of the shifts performed in these parameters when speakers vary their voice gender. Moreover, the focus of the present study on F0 and ΔF shifts does not exclude that adult speakers may have also spontaneously modified acoustic features other than the F0 and ΔF when varying their voice gender, and in turn, that such shifts may have also contributed to listeners' ratings. For example, Cartei and colleagues (2012) report that, while adults do not significantly vary their F0 variation (the acoustic correlate to intonation) when uttering vowels, speakers decrease (or increase) this parameter to sound more masculine (or feminine) in longer reading tasks. This is in line with perceptual and psychoacoustic studies showing that listeners rate men's voices as more monotonous than women's (Van Rie & Van Bezooijen, 1995; Wolfe, Ratusnik, Smith & Northrop, 1990). A few studies also report men's voices to be more creaky (Henton, 1989a), and less breathy (Klatt & Klatt, 1990) than women's, even though these perceptual differences are only weakly supported by the acoustic literature (Henton 1985, 1989a, 1989b; Klatt & Klatt, 1990; Simpson, 2009). The potential role of these parameters on listeners' ratings of gender remains a topic for future research.

Conclusions

Overall, our results show that, when pre-pubertal and adult speakers masculinised their voices (by lowering their F0 and ΔF), they were perceived as more masculine than when speaking normally. Similarly, when speakers feminised their voices (by raising F0 and ΔF) they were perceived as less masculine than when speaking normally. Taken together, the present observations lend further support to the "gender" hypothesis, by confirming the perceptual relevance of speakers' voice gestures to vary the expression of their gender.

Factors outside speakers' control, such as perceivers' internal state and social context, could also be critical in shaping both the use and interpretation of acoustic variation. For example, while the present study did not provide listeners with other sources of gender-individuating information apart from speakers' voices, in typical communicative situations individuals integrate contextual cues of their interlocutors (e.g. their sex, age, personality characterisations, behavioural traits, hairstyle, clothing style) when forming their impressions of them (Biernat, 1991; Fridell, Owen-Anderson, Johnson, Bradley, & Zucker, 2006; Kunda & Thagard, 1996; Mcdermid, Zucker, Bradley, & Maing, 1998). Thus, a more complete understanding of how each cue in different modes (e.g. auditory, visual) interacts with the others will do much to explicate resulting perceptions. In addition to contextual cues, perceivers are also likely to make inferences of a speaker's gender-related attributes from a combination of their own affective responses, and own knowledge of gender roles and stereotypes (e.g. how a man “should” sound (Devine, 1983; Smith & DeCoster, 1998)). As such, individual differences in listeners’ perceptions also need to be examined systematically.

References

Assmann, P. F., Dembling, S., & Nearey, T. M. (2006). Effects of frequency shifts on perceived naturalness and gender information in speech. In *INTERSPEECH*.

Bahrick, L. E., Netto, D., & Hernandez-Keif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*(5), 1263–1275.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231–238.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: a developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351–365.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.20)[Computer program]. Retrieved January 1, 2012.

Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2013). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology*, *32*(1), 100-106.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PloS One*, *7*(2), e31353.

Cartei, V., & Reby, D. (2013). Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children’s Voices. *PLoS ONE*, *8*(12), e81022.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men’s facial and vocal masculinity. *Evolution and Human Behavior*, *29*(4), 233–241.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., … Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, *49*(2), 215–222.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568.

Fridell, S. R., Owen-Anderson, A., Johnson, L. L., Bradley, S. J., & Zucker, K. J. (2006). The Playmate and Play Style Preferences Structured Interview: A comparison of children with gender identity disorder and controls. *Archives of Sexual Behavior*, *35*(6), 729–737.

Henton, C. G. (1985). A comparative study of phonetic sex-specific differences across languages (Doctoral dissertation, University of Oxford).

Henton, C. G. (1989a). Sociophonetic aspects of creaky voice. *The Journal of the Acoustical Society of America*, *86*, S26.

Henton, C. G. (1989b). Fact and fiction in the description of female and male pitch. *Language & Communication*, *9*(4), 299-311.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166.

Kewley-Port, D., Li, X., Zheng, Y., & Neel, A. T. (1996). Fundamental frequency effects on thresholds for vowel formant discrimination. *The Journal of the Acoustical Society of America*, *100*(4), 2462–2470.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820–857.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Mcdermid, S. A., Zucker, K. J., Bradley, S. J., & Maing, D. M. (1998). Effects of physical appearance on masculine trait ratings of boys and girls with gender identity disorder. *Archives of Sexual Behavior*, *27*(3), 253–267.

Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of Gender Stereotype Domains: Developmental and Gender Differences in Children. *Sex Roles*, *60*(11-12), 870–881.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech*, *50*(1), 125–142.

Ohara, Y. (1999). Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English. In *Wahrnehmung und Herstellung von Geschlecht* (pp. 105–116). Springer.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–2212.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988.

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The Queen's English: an alternative, biosocial hypothesis for the distinctive features of 'gay speech'. *Archives of Sexual Behavior*, *37*(1), 188–204.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.

Smith, E. R., & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology*, *74*(1), 21.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Van Rie, J., & Van Bezooijen, R. (1995). Perceptual characteristics of voice quality in dutch males and females from 9 to 85 years. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 290-293).

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging studya). *The Journal of the Acoustical Society of America*, *125*(3), 1666–1678.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental Sexual Dimorphism of the Oral and Pharyngeal Portions of the Vocal Tract: An Imaging Study. *Journal of Speech, Language, and Hearing Research*, *54*(4), 995–1010. doi:10.1044/1092-4388(2010/10-0097)

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, *110*, 464. doi:10.1121/1.1379087

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders*, *55*(1), 43.

Chapter 5: Speakers and Listeners apply the "Gender Code" to Gender Stereotypes

Summary

Having established the use of the "gender code" at both production and perception levels, the present chapter sets out to investigate its interpretation (by listeners) and application (by speakers) in different contexts, and in particular in relation to gender stereotypes. More specifically, the following question will be explored:

Question 6. How does the "gender code" interact with cultural stereotypes in the expression and perception of gender attributes and sexual orientation?

Study 7 investigates Question 6 at a perceptual level, by testing whether both pre-pubertal children and adults make stereotypical judgements about children's voices on the basis of children's stereotypical gender characterisations.

Summary of findings:

- Adults (as also shown in Study 2) and pre-pubertal children were sensitive to ΔF variation when making gendered characterisations of their peers
- Listeners spontaneously attributed more resonant (masculine) voices to stereotypically masculine descriptions of children (e.g. boy or girl playing with action toys), and less resonant (feminine) voices to stereotypically feminine descriptions of children (e.g. boy or girl playing with dolls)
- Associations between characters' voice and their gender-typed description varied according to speakers' and listeners' sex and age: e.g. boys preferentially assigned feminised voices to girl characters across conditions

Study 8 investigates Question 6 at a production level, by investigating whether actors playing homosexual roles feminise their voices by raising their F0 and ΔF, thus reproducing in the auditory dimension stereotypical notions that attribute feminine characteristics to male homosexuals.

Summary of findings:

- Actors’ F0 and ΔF were significantly higher when playing homosexual roles than when playing heterosexual roles or when being interviewed
- No significant differences in actors’ F0 and ΔF between heterosexual roles and interviews were found
- For homosexual roles, this “feminisation” effect was particularly pronounced in the comedy genre (in line with a more marked use of stereotypical characterisations in such genre)

Study 7:

Children and adults associate voice variation with gender-stereotypical characterisations of boys and girls

Note. Manuscript in preparation as: Cartei, V., Benarjee, R., Hardouin, L. & Reby, D. (2014). Children and adults associate voice variation with gender-stereotypical characterisations of boys and girls.

Abstract

Variation in the vocal cues signalling gender in adults (pitch and resonances) can lead to between and within-gender stereotyping, with lower-pitched, more resonant voices being attributed to more masculine individuals. Listeners can also discriminate the gender of pre-pubertal speakers from their voice, with boys speaking with lower resonances than girls. However, it is yet to be investigated whether resonance variation in children, as previously shown in adults, is associated with gender-stereotypical attributions. We re-synthesised the voices of seven-year old children (two boys and two girls) by artificially manipulating their formant spacing (ΔF) to a lower (masculinised voice), higher (feminised voice) and mid-point (prototypical voice) value within the pre-pubertal ΔF range. Using a cross-modal task, we then asked 15 pre-pubertal children and 18 adults to spontaneously associate prototypical pre-pubertal voices of boys and girls and their masculinised and feminised versions to boyish, mixed and girlish scenarios of boys' and girls' characters. We found that individuals spontaneously associate pre-pubertal gender-signalling cues (ΔF) to gender-stereotypical information about the child, although this ability varies with speakers' and listeners' sex and age. These results suggest that from childhood individuals spontaneously integrate stereotypical information in the auditory and visual domain to make stereotypical judgments about children's gender, thus highlighting the potentially important, and yet largely under-researched, role of the voice in gender stereotyping.

Introduction

The adult voice is strongly sexually dimorphic and listeners are able to reliably identify the gender of speakers from listening to their voice only. This ability is already present in six-month old infants (Miller, 1983), and reaches 98.8% accuracy in adults (Bachorowski & Owren, 1998). The two main cues to gender in adult voices are fundamental frequency and overall spacing of formant frequencies (Hillenbrand & Clark, 2009). On average, men's fundamental frequency (F0 (Titze, 1994)) is 50% lower than women's, giving male voices their lower pitch (Harries, Walker, Williams, Hawkins & Hughes, 1997). Men's formants are also lower than women's, resulting, on average, in 20% narrower spacing (ΔF) (Titze, 1994) and giving the male voices a "deeper" timbre (Baumann & Belin, 2010; Hollien, Green, & Massey, 1994). These acoustic dimorphisms are mainly the result of testosterone-driven changes to the vocal apparatus during male puberty, with men permanently developing ticker and longer vocal folds (inversely affecting F0) and longer vocal tracts (inversely affecting ΔF) than women (Titze, 1994). While F0 and ΔF encode adult speakers' gender, Cartei, Cowles and Reby (2012) have shown that adults spontaneously modify those parameters to vary the expression of their voice masculinity and femininity. Correspondingly, psychoacoustic studies have revealed that variation of these parameters is also used by listeners to stereotype adult speakers within the same gender. More specifically, listeners tend to attribute typically male attributes, such as confidence, dominance (Puts, Gaulin, & Verdolini, 2006), and masculinity (Assmann, Dembling, & Nearey, 2006; Ko, Judd, & Blair, 2006) to speakers with naturally occurring or artificially lowered F0 and ΔF, and typically female attributes, such as kindness, modesty and femininity to voices in which these parameters were naturally or artificially raised (Assmann, et al., 2006; Ko, Judd & Blair, 2006; Ko, Judd, & Stapel, 2009; Van Bezooijen 1995).

Voices of pre-pubertal children are also sexually dimorphic, despite no substantial differences in the size of their larynx and vocal tract cavities (Vorperian et al., 2009; Vorperian et al., 2011). Most acoustic studies show that, while boys and girls do not differ in F0, pre-pubertal boys speak with lower formants (Fi) and narrower spacing (ΔF) than girls (Bennett, 1981; Cartei, Cowles, Banerjee & Reby, 2013; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001), leading some authors to suggest early acquisition of gender-linked ways of speaking (Lee et al., 1999;

Sachs, Lieberman, & Erickson, 1973). Correspondingly, voice gender differences in children's voices are attended to by listeners from early childhood: seven-month old infants have been found to match the faces and voices of unknown children (Bahrick, Netto, & Hernandez-Keif, 1998) and adults are capable of identifying speakers' gender from the voice of children as young as four with good level of accuracy (from 66% to 81%, Karlsson, 1989; Perry et al., 2001; Sachs et al., 1973).

Cartei and colleagues (2013) have also shown that children spontaneously alter formant spacing ΔF, to sound 'like a boy' and 'like a girl', indicating that this parameter is likely to function as cue to masculinity and femininity in children's voices, as previously established in adults (Feinberg 2008; Fraccaro et al., 2010; Pisanski & Rendall, 2011). While no studies have explored whether children listeners link voice variation to gender-related characteristics, manipulations of ΔF within the children's natural range have been found to affect adult listeners' judgments of vocal masculinity of a child speaker (e.g. voices artificially lowered in ΔF received higher vocal masculinity ratings (Cartei & Reby, 2013)). Moreover, research using visual stimuli suggests that since childhood individuals tend to over-generalise sex signalling cues to stereotype their peers. More specifically, while younger children (up to about five years of age) make strong gender stereotypical assumptions which are solely based on the gender label (boy or girl), older children (from seven years old onwards) are able to take into account the child's gender as well as gender-typed characteristics and preferences (Banerjee et al., 2000; Biernat, 1991; Martin, Ruble, & Szkrybalo, 2002). For example, when presented with a masculine or feminine characteristic about a child of unknown gender, seven and eight year olds can make stereotypical predictions in multiple dimensions (e.g. physical appearance, interests) on the basis of that characteristic alone (Martin, Wood & Little, 1990). Moreover, older children are more likely to attribute stereotypical characteristics of the opposite gender to pre-pubertal children displaying counter-stereotypical behaviour (Lobel & Menashri, 1981), choice of toys (Martin, Eisenbud, & Rose, 1995), activities (Banerjee & Lintern 2000) and traits (Carter & McCloskey, 1984).

In the present study we use a cross-modal task to explore children's spontaneous associations of voices with stereotypical and counter-stereotypical characterisations of children's gender identity (as represented by textual and pictorial descriptions of friendship preferences, gendered activities and toy preferences), and compare those results with those obtained by having the same task performed by adults. As we do not explicitly ask listeners to rate these voices as more or less masculine, their associations are spontaneous and therefore likely to reflect how people interpret gender-related vocal cues in the real world.

Hypotheses

The goal of this study is to determine the extent to which children and adult listeners match voices of pre-pubertal children manipulated in ΔF with gender stereotypical descriptions. In particular, we propose that:

(i) masculinised male voices (lowered ΔF) will be associated with stereotypical descriptions of boy characters, e.g. boys exhibiting male friends and masculine toy and activity preferences

(ii) feminised male voices (raised ΔF) to counter-stereotypical descriptions of boy characters, e.g. boys who favour playing with girls and prefer feminine toys and activities

(iii) prototypical male voices (ΔF re-synthesised to a middle value between the two gendered voices) to boy characters who play with mixed groups and exhibit gender-neutral activity and toy preferences

In a corresponding and opposite pattern, we further propose that:

(i) masculinised female voices (lowered ΔF) will be associated with counter-stereotypical descriptions of girl characters, e.g. girls exhibiting male friends and masculine toy and activity preferences

(ii) feminised female voices (raised ΔF) to stereotypical descriptions of girls characters, e.g. girls who favour playing with girls and prefer feminine toys and activities

(iii) prototypical female voices (ΔF re-synthesised to a middle value between the two gendered voices) to girl characters who play with mixed groups and exhibit gender-neutral activity and toy preferences

We also evaluated the possibility that adults would be more sophisticated than children when integrating auditory information with other information about the characters, given the well established development of gender-related cognition (e.g. stereotype knowledge and identity (Martin et al., 1994; Biernat, 1991)) and audio-visual cue integration (Burri & Gory, 2012).

Methods

Acoustic measurements

Speech utterances were recorded using a *Shure SM94* microphone connected to a *Tascam DR07mkII* handheld recorder in a Sussex primary school, as part of a study of gender expression in children's voices. During those recordings children were asked to read six words out loud: *hat*, *duck, bed*, *feet*, *book, boot*. The words of two seven-year old males and two seven-year old females were selected and concatenated with an interval of 50 ms silence. Samples were then scaled in intensity to a 65 dB level. Fundamental frequency (*F*0) values and the frequency of the first four formants (*F*1–*F*4) were obtained using PRAAT v.5.20 (Boersma & Weenink, 2011). Mean F0 values were obtained using PRAAT's pitch-tracking algorithm "To pitch" over the entire sequence with a range setting of 100–500 Hz and time step 0.01s. The mean values of the first four formants (F1–F4) were measured from the list of words using the LPC Burg algorithm in PRAAT. The algorithm's parameters were initially set as number of formants 5, max formant 6000–6600 Hz, dynamic range 30dB and window length of 0.025s, and adjusted manually to visually obtain the best fitting prediction (one where the predicted formants are superimposed as much as possible onto the observed formants in the spectrogram). Formant spacing (ΔF) was calculated from F1–F4 using the procedure specified by Reby and McComb (2003). According to this model, the vocal tract can be approximated as a straight uniform tube closed at the glottis and opened at the lips. Under such model, F_i are expressed as:

$$(1) \quad F_i = \frac{2(i-1)c}{4aVTL}$$

Where *i* is the formant number, *c* is the speed of sound in a mammal vocal tract (350m/s), *aVTL* is the apparent vocal tract length and *F*i is the frequency of *i*th formant. From (1) and, it follows that:

$$(2) \Delta F = F_{i+1} - F_i = \frac{c}{2aVTL}$$

By replacing 2c/aVTL with ΔF in equation (1), ΔF can be derived as the slope of a regression model with the observed Fi values (y-axis) plotted against the expected formant positions:

$$F_i = \frac{(2i-1)}{2}\Delta F \quad (3)$$

The above formula is a good estimate of ΔF because, while individual formants are affected by the shape as well as the length of the vocal tract required to express the different sounds, the formant spacing is an average of adjacent formant differences, and thus provides an overall estimate of spectral dispersion which is less sensitive to such deviations. Additionally, as ΔF is determined by, and inversely correlated to, the length of the vocal tract of the speaker (Titze, 1994), it follows from (2) that the apparent vocal tract length can be estimated as aVTL = c/2(ΔF). Apparent Vocal Tract Length (aVTL) was calculated because it can be expressed in cm and therefore gives a better illustration of the scale of the manipulations performed than ΔF (which is expressed in Hz). The resulting mean values of F0, F1–F4 and ΔF (Table 5.7.1) agree well with those reported by previous acoustic studies (Lee et al., 1999; Cartei et al., 2013). Apparent VTL values (Table 5.7.1) are also comparable to those estimated by acoustic studies (Lee et al., 1999) and to typical VTL values reported by MRI data (Fitch & Giedd, 1999).

Table 5.7.1

Original acoustic values (F0, F1–F4, ΔF, aVTL) for girls and boys' exemplars

Voice stimuli	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	ΔF (Hz)	aVTL (cm)
Female 1	252	983.0	2326.8	3726.2	4618.0	1402.9	12.5
Female 2	253	1035.3	2334.3	3719.8	4708.1	1418.9	12.3
Male 1	248	1146.1	2203.1	3546.1	4592.4	1372.2	12.8
Male 2	250	976.6	2248.3	3419.1	4724.7	1378.3	12.7

Acoustic manipulations

All manipulations were carried out using the PSOLA algorithm in PRAAT. This method allows for ΔF (and thus aVTL) manipulation, while preserving F0 and duration. The original samples were re-synthesised to approximate the target ΔF of 1300Hz (aVTL of 13.5cm) in boys and 1400Hz (aVTL of 12.5cm) in girls for the prototypical condition. The k-factors required to change the apparent vocal tract lengths of our exemplars to these values were obtained by dividing either 1300Hz or 1400Hz by the original measured ΔF for each exemplar.

The target values for the masculinised and feminised conditions were chosen in line with the end points of the pubertal (5–11 year old) children's ΔF range, estimated from published formant frequency values (Lee et al., 1999; Cartei et al., 2013) as well as the perceived range of the two sexes from psychoacoustic studies (Cartei & Reby, 2013) and corresponding about 88% or 112% of the prototypical condition. For the masculinised voices the entire sound spectrum was scaled down (thus lowering all formants and narrowing their spacing) in order to approximate the target ΔF of 1160Hz in boys and 1250Hz in girls, corresponding to aVTLs of 15cm in boys and 14cm in girls. For the feminised voices ΔF was raised to 1450Hz (aVTL of 12cm) in boys and 1580Hz (aVTL of 11cm) in girls. Again, the k-factors required to change the apparent vocal tract lengths of our exemplars to these values were obtained by dividing either 1160Hz or 1450Hz by the original measured ΔF for each exemplar. In summary we generated 12 audio stimuli from the word lists spoken by the four children.

Stimuli presentation

A cross-modal task was used to assess whether children could make stereotypical predictions about a child character's voice when given both sex and gender-linked information about playmates, toys and activities. Boy and girl characters were presented in separate Microsoft PowerPoint presentations. Each slide in the presentation depicted the child character in one of the three (boyish, girlish, mixed) scenarios with the three audio stimuli (masculinised, prototypical, feminised) from the same exemplar. The visual information in the "boyish", "girlish" and "mixed" scenarios was presented using gender-stereotypical, gender-neutral and counter-stereotypical text and cartoon-style pictorial descriptions of friendship, toy and activity preferences. This type of visual information allowed us to overcome possible issues with vocabulary use, as young children may not have the necessary understanding of gender-related (e.g. "masculine or boyish" or "feminine or girlish") labels (Martin, 1990). Moreover, to account for differential reading abilities, an audio recording of the description accompanied each slide. For example, for boys' characters the boyish scenario stated that the child had male friends and masculine interests (e.g. "this is a boy called Phil.

Phil really likes to play with train sets and action toys. He likes playing football with the boys in his class"). In the girlish scenario the boy had girl friends and feminine interests (e.g. " this is a boy called Mark. Mark really likes to play with dolls and puzzles. He likes playing dressing-up games with the girls in his class"). In the mixed scenario the child had friends of both genders and gender-neutral interests (e.g. "this is a boy called Danny. Danny really likes to play with farm animals and board games. He likes playing with the boys and girls in his class"). Below the text description, the slide contained three visual scenarios of the toys, followed by the question "Which voice is [boy's name]'s?" and three buttons labelled "1", "2", "3". Each button played one of the three re-synthesised voices (masculinised, prototypical, feminised) from the same exemplar. Thus the three scenarios (girlish, boyish and mixed) were shown twice, once per exemplar, for a total of six slides per presentation. The slide order was alternated (e.g. if the first slide contained voices from exemplar 1, the second slide would contain voices from exemplar 2) and another PowerPoint presentation was created with the association between characters and exemplars reverted. Participants were alternatively shown one of the two presentations, to minimise order effects on their choices. The PowerPoint slides containing girls' characters were created using the same methodology and presentations of the boys' and girls' characters were counterbalanced. The choice of gendered and gender-neutral toys was based on previous research revealing that children show no clear gender stereotypical preferences for toys such as farm animals and play doh (Goble, Martin, Hanish, & Fabes, 2012; Miller, 1987; Stagnitti, Rodger, & Clarke, 1996). However, girls tend to prefer soft toys, dolls and domestic furniture (Marcon & Freeman, 1996; Rubles, Martin, & Berenbaum, 2006; Smith & Daglish, 1977), and boys prefer transportation toys and action figures (Jones & Glenn, 1991; Marcon & Freeman, 1996; Rubles et al., 2006).

Participants

Children participants were 10 boys and eight girls, aged seven to nine (*M* = 7.82, *SD* = 1.1), who were recruited from a primary school in West Sussex and from Brighton, East Sussex. Adult participants were eight male and 10 female students, aged 20 to 28 (*M* = 22.34, *SD* = 2.2), who were recruited from the University of Sussex.

Participants' consent and participants' guardians (for children) consent were obtained prior the experiment.

Procedure

Participants were individually tested in a quiet room at a school (children) or at the University of Sussex (children and adults). Participants were sat in front of a laptop computer and wore *Dynamode dh-660mv* headsets. Firstly, they were made to listen to a neutral sound in order to adjust the sound volume, which was set to a comfortable level (default value of 65%). Three adults and six children changed the volume by moving the lever on the headphones. Next, participants were showed a PowerPoint presentation with boy-only characters, followed by a presentation with girl-only characters, in alternate order. For each slide, participants viewed and listened to the character's description. Next, they listened to the three voices through their headsets, one at the time, by clicking on the buttons labelled "1", "2" and "3". Adults decided which of the three voices belonged to the character depicted in the slide by circling the number of the voice on a response sheet. Children voiced their judgments, and the experimenter marked their answers on the response sheet. Once the choice was made, participants moved on to the next slide by clicking on the "next" arrow button.

Statistical Analyses

We ran two separate Generalised Linear Mixed Models by character's sex (boy and girl characters) to assess the effect of age group (children, adults), participant sex (male, female) and scenario (boyish, mixed and girlish) on the choice of voice (masculinised, prototypical, feminised) attributed to the characters. All analyses were run using the GLMM procedure in SPSS v.20.0 for Windows Vista. Initially, we fit full models where the choice of voice was specified as the response variable ("target" in SPSS, with default value set to prototypical voice). Group, participant sex, scenario and their first and second order interactions were specified as fixed factors ("fixed effects" in SPSS), while participant identity and character number (boys 1 and 2, girls 3 and 4) were entered into the model as random factors ("random effects" in SPSS). For all models, we started from the global model (including all explanatory variables, first and second order interactions) and compared it with sub models from which we sequentially deleted non-significant terms until all of the remaining terms were significant. We used the corrected Akaike's information criterion (AICc) for final sub model comparisons and selection. We considered two models to be significantly different when the difference between their AICcs was greater than 2 (Burnham & Anderson, 2002).

Results

The selected two models showed a significant main effect of *scenario* for both boys, $F(4, 196) = 11.19$, $p < .001$, and girls characters, $F(4, 178) = 3.61$, $p = .007$. The interaction effect of *scenario*group* was also significant, revealing that the choice differed according to the age of participants (group: adults or children) and scenario for both boys, $F(4, 196) = 2.64$, $p = .035$, and girls characters, $F(4, 178) = 2.61$, $p = .037$. In addition, the girl character model revealed a significant interaction effect of *group*participant_sex*, $F(2, 178) = 5.94$, $p = .003$. To further investigate contrasts between the three levels of our response variable (choice of voice) in light of the above interactions, we separated our data according to group, resulting in four sub-models (boy characters/adults, boy characters/children, girl characters/adults, girl characters/children), with *scenario* as the only fixed factor in boy characters models, and *participant_sex* and *scenario* as fixed factors for girl characters models (due to the significance of the interaction between *scenario* and *group*). Between-scenario differences in the voices chosen for each of the four models were explored via pair-wise contrasts between each voice pair (given in Table 5.7.2) from GLMM separated by characters' sex: the two sub-models for boy characters had scenario as fixed factor (and the same random factor structure), while the two sub-models for girl characters had *scenario*, *participant_sex* and their interaction as fixed factors.

To explore within-scenario differences in the voices chosen in each of the four models, contrasts from repeated-measures ANOVAs were run for each scenario with voice (masculinised, prototypical, feminised) as a three-level within-subject factor and illustrated in in Figure 5.7.1 (horizontal bars). Standardised percentages representing how often a voice was chosen within each scenario were obtained using Cross-tabs procedures and illustrated in Figure 5.7.1 (vertical bars).

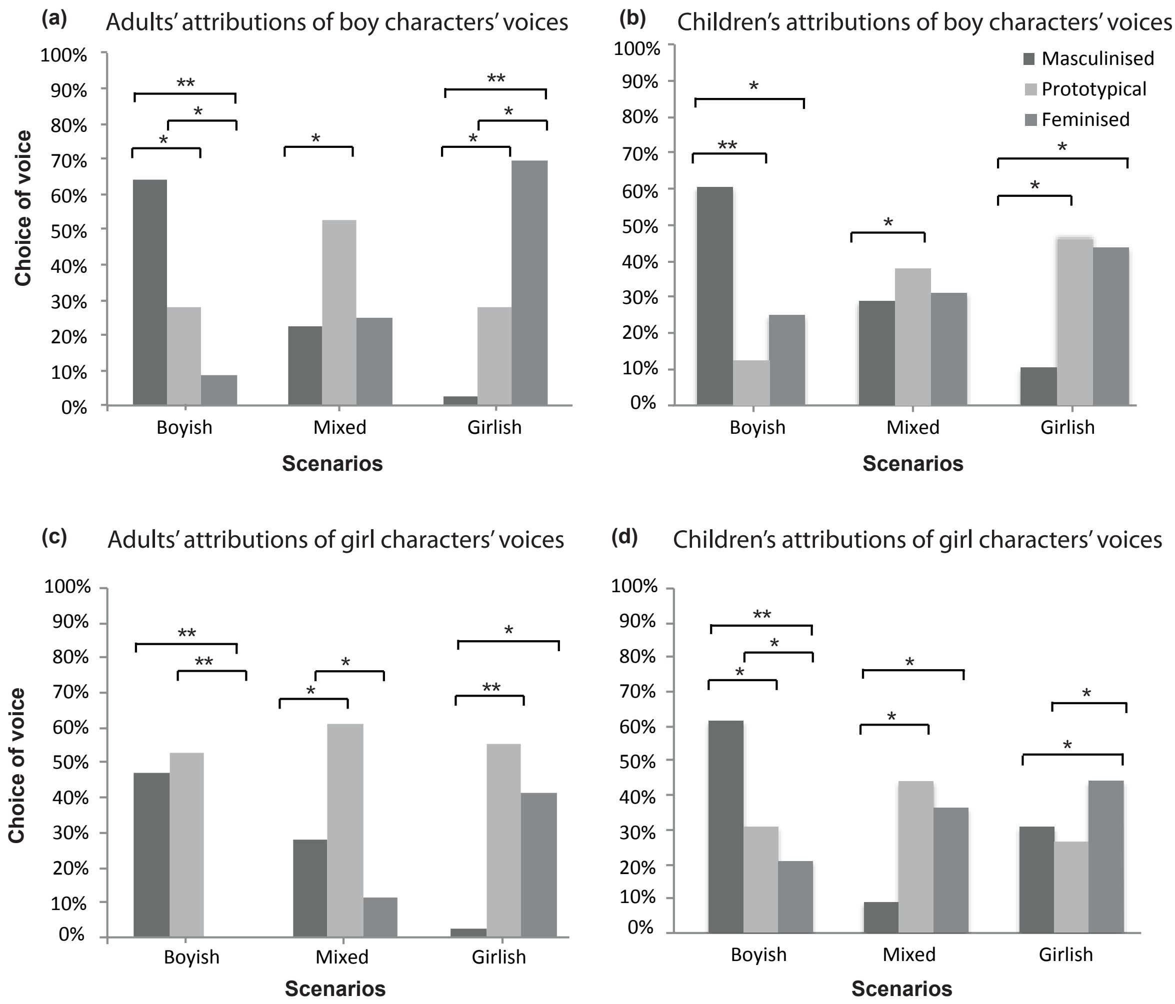

Figure 5.7.1. Voice attributions for boy and girl characters by adults (a, c) and child (b, d) raters according to the presented scenarios. Black bars represent the masculinised voices, light grey bars represent the prototypical voices and dark grey bars represent the feminised voices. Contrasts' significance from repeated-measures ANOVAs (voice as within-subject factor) for adult and child raters in boy and girl characters and within each scenario, are represented by horizontal bars, with p ** <.001; * significant at p < .05.

Table 5.7.2

Contrasts resulting from the sub-models Boy character / Adult participants, Boy character /Child participants, Girl character / Adult participants and Girl character / Child participants

		Scenarios a.Girlish vs. Mixed		b.Boyish vs. Mixed		c.Boyish vs. Girlish	
	Boy character / Adult participant	Estimate	p	Estimate	p	Estimate	p
1	Prototypical vs. Masculinised	-1.438	.206	1.698	**.003**	**3.136**	**.006**
2	Feminised vs. Masculinised	**-3.123**	**.007**	2.186	**.006**	**5.309**	**< .001**
3	Prototypical vs. Feminised	**2.291**	**.001**	-.722	.391	**-3.013**	**.001**
	Boy character / Child participant						
4	Prototypical vs. Masculinised	-1.204	.081	**1.854**	**.004**	**3.058**	**< .001**
5	Feminised vs. Masculinised	1.347	.063	.978	.108	**2.336**	**.001**
6	Prototypical vs. Feminised	.154	.777	.875	.215	.721	.286
	Girl character / Adult participant						
7	Prototypical vs. Masculinised	**-2.377**	**.038**	.813	.136	**3.117**	**.006**
8	Feminised v.s Masculinised	**3.88**	**.002**	2.014	.098	**5.894**	**< .001**
9	Prototypical vs. Feminised	**1.875**	**.013**	-1.28	.292	**2.755**	**.015**
	Girl character / Child participant						
10	Prototypical vs. Masculinised	**1.776**	**.048**	**2.377**	**.006**	.601	.334
11	Feminised vs. Masculinised	-1.027	.260	**-2.665**	**.004**	**1.637**	**.017**
12	Prototypical vs .Feminised	.728	.182	-.323	.585	-1.051	.087

Note: statistically significant contrasts are shown in bold

Boy characters

As predicted, within each scenario, adults attributed the congruent voice to the character's portrayed "gender" (Figure 5.7.1a): subjects presented with the boyish scenarios attributed masculinised voices significantly more than prototypical and feminised voices to the character. Similarly, subjects presented with the girlish scenarios attributed feminised voices significantly more than masculinised or prototypical voices to the character. Finally subjects presented with the mixed scenarios attributed prototypical voices more than masculinised voices (significant trend) and feminised voices (non-significant trend, p = .073). As a consequence, the type of visual scenario (boyish, mixed or girlish) had a significant effect on which variant of resynthesised boys' voices (masculinised, prototypical and feminised) was attributed to the boy character by adult raters, $F(4, 102) = 9.29$, $p < .001$ (Figure 5.7.1a): prototypical and feminised voices were similarly (less) attributed in the boyish scenarios (Table 5.7.2, contrast 3b), while prototypical and masculinised voices were similarly (less) attributed in the girlish scenarios (Table 5.7.2, contrast 1a).

As with adults, children chose the masculinised voices significantly more than the other two voices in the boyish scenarios and the prototypical voices more than the boyish voices in the mixed scenarios (Figure 5.7.1b). Also similarly to adults, in the girlish scenarios children chose the feminised voices significantly more than the masculinised voices. These attributions were reflected in between scenario differences with scenario having a significant main effect on choice of voice by children raters, $F(4, 102) = 4.99$, $p = .001$ (Figure 5.7.1b). As with adults, masculinised voices were preferably attributed when facing the boyish scenarios, compared to the mixed and girlish scenarios (Table 5.7.2, contrasts 4a, 4b, 4c, 5c), while feminised and prototypical voices were similarly more attributed than masculinised voices in the girlish scenario (Table 5.7.2, contrasts 4a, 5a, 4c, 5c).

Girl Characters

Adults' voice attributions of girl characters were also largely congruent with the gender-typed portrayals (Figure 5.7.1c): as expected, in the boyish scenarios adults chose the masculinised voices significantly more than the feminised voices (but not more than the prototypical voices). Similarly, in the girlish scenarios, the feminised voices were significantly more chosen than the masculinised voices (but not more than the prototypical voices). In the mixed scenarios, adults chose the prototypical voices significantly more than the other two. These differences were reflected in differences between scenarios (Figure 5.7.1c). GLMM revealed that scenario had a significant main effect on choice of voice by adult traters, $F(4, 102) = 3.8$, $p = .006$. Masculinised voices were less attributed in the girlish scenarios compared to the other two (Table 5.7.2 - 7a, 7c, 8a, 8c), while feminised voices were less attributed in the boyish and mixed scenarios compared to the girlish scenarios (Table 5.7.2, contrasts 8a, 9a, 8c, 9c). The sex of adult participants was not significant, $F(2, 100) = 2.2$, $p = .114$.

In line with our hypotheses, children attributed masculinised voices significantly more than feminised or prototypical voices in the boyish scenarios, and the feminised voices significantly more than the other two in the girlish scenarios (Figure 5.7.1d). Children also significantly chose prototypical voices more than masculinised voices (but not more than feminised voices) in the mixed scenarios. These attributions were reflected in between-scenario differences with scenario having a significant main effect on the attribution of voice variants to characters, $F(4, 100) = 3.91$, $p = .013$ (Figure 5.7.1d): prototypical and feminised voices were similarly (less) attributed in the boyish scenarios (Table 5.7.2, contrast 12b), while prototypical and masculinised voices were similarly (less) attributed in the girlish scenarios (Table 5.7.2, contrast 10c).

In contrast with adults, the choice of voice attributed to girl characters was significantly affected by the sex of children participants, $F(2, 100) = 6.5$, $p = .002$: GLMM contrasts (Table 5.7.3) and cross-tabs revealed that, across scenarios, boys were less likely to choose masculinised (boys: 39.1%, girls: 60.9%) and prototypical (boys: 45.7%, girls: 54.3%) voices than girls, but more likely to choose feminised voices (76.9%) than girls (23.1%).

Table 5.7.3

Contrasts resulting from male children vs. female children raters across scenarios

Choice of Voice	Estimate	p
Prototypical vs. Masculinised	.383	.490
Feminised vs. Masculinised	**1.990**	**.002**
Prototypical vs. Feminised	**-1.460**	**.004**

Note. statistically significant contrasts are shown in bold

Discussion

We hypothesised that listeners would attribute masculinised voices (whose ΔF was artificially lowered) to masculinised characterisations of children, feminised voices (whose ΔF was artificially raised) to feminised characterisations and prototypical voices (whose ΔF was re-synthesised to a middle value between the two gendered voices) to gender-neutral characterisations. In line with our hypotheses we found that individuals associate within-gender variations in speakers with gender-typed information about their choice of playmates, toys and activities. We also tested the possibility that adults would perform more sophisticated attributions than children, but found no clear evidence of improvement in task performance with age, though age-related differences were observed. The implications of these results on the potential role of voice variation in gender stereotyping are discussed below, in the light of the observed age and gender differences.

Both men and women made stereotypical judgments when told about an unknown boy character with masculine or feminine traits (toys, activities, sex of playmates), thus showing that adults used stereotypical information to make their inferences. This pattern of results is compatible with observations from a previous psychoacoustic study (Cartei & Reby, 2013), in which adults were asked to rate children's voices resynthesised in ΔF along a gender continuum from "masculine boy" to "feminine girl". The study showed that adults were more likely to rate boys' voices whose ΔFs averaged at about 1160Hz (corresponding to the present masculinised variant, aVTL of 15 cm) as belonging to "masculine boys" or "boys", while those with ΔFs around 1450Hz (corresponding to the present feminised variant, aVTL of 12 cm) were more likely to be perceived as belonging to children of unknown gender or "masculine girls".

When assessing boy characters, children also took into account stereotypical information to the extent that they associated boys' masculinised voices to boyish scenarios and did not choose those voices for the girlish scenarios. However, they did not preferably assign boys' feminised voices to the girlish scenario and in fact, boys' feminised and prototypical voices were similarly chosen in all scenarios. One possibility is that children, unlike adults, were not able to perceive differences between prototypical and feminised versions of the same vocal stimulus. Although the acoustic difference between any voice pair was the same (12%) and equal to twice the just noticeable difference (JND) in ΔF (Pisanski & Rendall, 2011; Puts Hodges, Cárdenas, & Gaulin, 2007), frequency differences become perceptually smaller with increasing frequency (Madisetti, 2009). Consequently, the prototypical and feminised voices could have been perceived as closer together than prototypical and masculinised voices. Future studies could investigate whether varying ΔF by equally small amounts along the natural range of children's voices would elicit consistent effects on gender-based assessments despite the logarithmic nature of frequency perception. Some children may have also deliberately avoided the feminine voice for the boy character even though he displayed feminine interests. Research shows that individuals, and especially children, negatively evaluate femininity traits in boys who adopt cross-gender characteristics more so than in girls, and at the same time are less likely to attribute negative qualities to their own gender (Martin et al., 1990; Martin et al., 1995).

Adult and children listeners' voice attributions voices to girl characters were also broadly comparable. More specifically, while adults rarely chose the incongruent voice for gender-typed characterisations (e.g. never choosing feminised voices for girls engaging in boyish behaviour, and hardly choosing masculinised voices for girls engaging in girlish behaviour), children (though not adults) preferably assigned masculinised voices to boyish scenarios and feminised voices to girlish scenarios. One possibility for the observed age-related discrepancies is that the relative shifts in ΔF for the feminised and masculinised variants were sufficient to elicit the expected gendered attributions in children, but not in adults, suggesting that children may be more attuned than adults to ΔF differences in girls' voices. Interestingly, Cartei and Reby (2013)'s investigation of adults' sensitivity to gender-related ΔF variation found that girls' voices with ΔF of around 1550 Hz (corresponding to aVTL of 11.4cm) were rated by adult listeners as belonging to a "girl" rather than to a "feminine girl". Though a higher ΔF value was used for the feminised voices in the present study (1580Hz, corresponding to an aVTL of 11cm), adult listeners still failed to associate them to stereotypical characterisations of girls. Future work is needed to study the effect of ΔF variation on gender attributions in children and compare these with previous research in adults (Cartei & Reby, 2013), in order to explore whether ΔF differentially cues for gender characteristics according to listeners' and speakers' sex and age.

Furthermore, differences in girls' attributions may also reflect developmental and gender-specific differences in stereotype rigidity. For example, adults tend to treat activities and behaviours labelled as 'masculine' or 'feminine' as appropriate for females, while children, and especially boys, tend to view girls as more stereotypically feminine (Feinman, 1981; Archer, 1992). The latter result may account for the present observation that, across scenarios, boys preferred feminised voices for girl characters. Taken together these results suggest that, while counter-stereotypical characterisations of girls may have been sufficient to trigger a non-prototypical voice attribution in children, especially in boys, more extreme counter-stereotypical characteristics may be needed to trigger a non-prototypical voice attribution in adults.

While much research has so far focused on children's development of gender stereotypes in the visual domain, the role of the voice in gender stereotyping remains largely unexplored. The present study shows that adults and children listeners spontaneously link resonance (ΔF) variation in children's voices to gender stereotypical characterisations. Our observations provide the first evidence that children are at least partially sensitive to ΔF cues when making judgments of gender-related attributes in their peers, while confirming this sensitivity in adult listeners (Cartei & Reby, 2013). While we forced participants to make stereotypical judgments, future studies will need to explore how continuous variation of children's resonances along the gender continuum affects children's masculinity and femininity ratings.

The present study also revealed that the auditory thresholds at which gendered characterisations trigger gender-typed voice attributions varies with gender of speaker, gender of listener and listener's age. More research is needed to investigate the extent to which these differences are linked to perceptual ability as well as to developmental and gender-specific differences in the use of gender-stereotyped knowledge. For example, future studies could investigate whether voice features other than ΔF play a role in individuals' impressions of children's gendered identities. In particular, studies could investigate the role of F0, which is sexually dimorphic in adults, but not in children (Titze, 1994), in attributions of gender-related traits to children's voices. The real world relevance of these psychoacoustic investigations should also be confirmed with perceptual studies examining the co-variation of acoustic cues in natural voices with children's perceived gender-related attributes.

Finally, investigating when and how the ability to make associations between auditory variation and gender stereotypes develops, taking into account age, gender and individual preferences, may provide further cues to explain how individuals learn and use gender-related information when judging others and thus contribute to our understanding of how gender roles are developed and maintained.

References

Assmann, P. F., Dembling, S., & Nearey, T. M. (2006). Effects of frequency shifts on perceived naturalness and gender information in speech. In *INTERSPEECH*.

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America*, *106*(2), 1054–1063.

Bahrick, L. E., Netto, D., & Hernandez-Keif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*(5), 1263–1275.

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408.

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research Psychologische Forschung*, *74*(1), 110–120.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231–238.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: a developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351–365.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.20)[Computer program]. Retrieved January 1, 2012.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Burr, D., & Gori, M. (2012). Multisensory Integration Develops Late in Humans. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press.

Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2013). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology*, n/a–n/a.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PloS One*, *7*(2), e31353.

Cartei, V., & Reby, D. (2013). Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children's Voices. *PLoS ONE*, *8*(12), e81022.

Carter, D. B., & McCloskey, L. A. (1984). Peers and the Maintenance of Sex-Typed

Behavior: The Development of Children's Conceptions of Cross-Gender Behavior in Their Peers. *Social Cognition*, *2*(4), 294–314.

Feinman, S. (1981). Why is cross-sex-role behavior more approved for girls than for boys? A status characteristic approach. *Sex Roles*, *7*(3), 289–300.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior*, *29*(4), 233–241.

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology*, *8*(3).

Goble, P., Martin, C., Hanish, L., & Fabes, R. (2012). Children's Gender-Typed Activity Choices Across Preschool Social Contexts. *Sex Roles*, *67*(7), 435–451.

Harries M. L. L., Walker J. M., Williams D. M., Hawkins S., Hughes I. A. (1997) Changes in the male voice at puberty. *Archives of Disease in Childhood, 77*, 445–447.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America*, *96*(5), 2646–2654.

Karlsson, I. (1989). Dynamic voice source parameters in a female voice. *Speech Transmission Laboratory-Quarterly Progress and Status Report, Royal Technical Institute, Stockholm*, *1*, 75–77.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the Voice Reveals: Within- and Between-Category Stereotyping on the Basis of Voice. *Personality and Social Psychology Bulletin*, *32*(6), 806–819.

Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin*, *35*(2), 198–211.

Jones, A., & Glenn, S. M. (1991). Gender differences in pretend play in a primary school group. *Early Child Development and Care*, *77*(1), 127–135.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Lobel, T. E., & Menashri, J. (1993). Relations of conceptions of gender-role transgressions and gender constancy to gender-typed toy preferences. *Developmental Psychology*, *29*(1), 150–155.

Madisetti, V. (2009). *Video, Speech, and Audio Signal Processing and Associated Standards*. CRC Press.

Marcon, R. A., & Freeman, G. (1996). Linking gender-related toy preferences to social structure: Changes in children's let- ters to Santa since 1978. *Journal of Psychological Practice, 2*, 1–10.

Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's Gender-Based Reasoning about Toys. *Child Development*, *66*(5), 1453–1471. doi:10.1111/j.1467-8624.1995.tb00945.x

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*(6), 903.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development*, *61*(6), 1891–1904.

Miller, C. L. (1987). Qualitative differences among gender- stereotyped toys: Implications for cognitive and social development in girls and boys. *Sex Roles*, *16,* 473–487.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–2212.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N.

Eisenberg, W. Damon, & R. Lerner (Eds.), *Handbook of child psychology*, 6th ed.: Vol 3 (pp. 858-932). Hoboken: Wiley.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Smith, P. K., & Daglish, L. (1977). Sex differences in parent and infant behavior in the home. *Child Development*, 1250–1254.

Stagnitti, K., Rodger, S., & Clarke, J. (1997). Determining gender-neutral toys for assessment of preschool children's imaginative play. *Australian Occupational Therapy Journal*, *44*(3), 119–131.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265.

Study 8:

Acting Gay: Male Actors Shift the Frequency Components of Their Voices Towards Female Values When Playing Homosexual Characters

Note. Study published as: Cartei, V., & Reby, D. (2012). Acting Gay: Male Actors Shift the Frequency Components of Their Voices Towards Female Values When Playing Homosexual Characters. *Journal of Nonverbal Behavior*, *36*(1), 79–93.

Abstract

The purpose of this study was to investigate whether actors playing homosexual male characters in North-American television shows speak with a feminised voice, thus following longstanding stereotypes that attribute feminine characteristics to male homosexuals. We predicted that when playing homosexual characters, actors would raise the frequency components of their voice towards more stereotypically feminine values. This study compares fundamental frequency (F0) and formant frequencies (Fi) parameters in the speech of fifteen actors playing homosexual and heterosexual characters in North-American television shows. Our results reveal that the voices of actors playing homosexual male characters are characterised by a raised F0 (corresponding to a higher pitch), and raised formant frequencies (corresponding to a less baritone timbre), approaching values typical of female voices. Besides providing further evidence of the existence of an ‘‘effeminacy’’ stereotype in portraying male homosexuals in the media, these results show that actors perform pitch and vocal tract length adjustments in order to alter their perceived sexual orientation, emphasising the role of these frequency components in the behavioural expression of gender attributes in the human voice.

Introduction

The portrayal of male homosexuals in films and television often follows an ‘‘effeminacy stereotype’’ (Kite and Deaux, 1987; Madon, 1997), which attributes feminine connotations to adult male homosexuals. Whilst acknowledging that stereotypes affect multiple dimensions of behaviour (Blashill & Powlishta, 2009), previous literature has focused on the feminisation of homosexual characters' mannerisms and lifestyles (Battles & Hilton-Morrow, 2002; Blashill & Powlishta, 2009; Chung, 2007; Linneman, 2008; Raley & Lucas 2006; Staats, 1978), but has overlooked the characters' voices. The present study investigates whether the voices of homosexual characters are feminised, that is, whether actors playing such roles modify (either consciously or unconsciously) their habitual voice towards values characteristic of heterosexual female voices.

According to the source-filter theory (Fant, 1960), the production of the human voice is characterised by two successive and independent stages. First the glottal wave is generated in the larynx (the ‘‘source’’) by periodic vibration of the vocal folds. This wave is a complex periodic signal with a fundamental frequency, or F0 (equal to the rate of glottal vibration, and responsible for the perceived ‘‘pitch’’ of the voice), and its integer multiple frequencies, the harmonics. As the glottal wave propagates from the larynx to the lips, the vocal tract acts as a filter and selectively amplifies or dampens frequencies, producing spectral peaks called formant frequencies (which affect the perceived ‘‘timbre’’ of the voice). While the source- and filter-related components can vary independently, they are constrained by the dimensions of the vocal apparatus (Fant, 1960). Compared to women, men speak with a lower F0 (Hollien et al., 1994; Lee et al., 1999; Perry et al., 2001; Rendall et al., 2005; Wolfe et al., 1990), giving them a lower pitch, and also with lower formant frequencies (Busby & Plant, 1995; Lee et al., 1999; Perry et al. 2001; Rendall et al., 2005), giving them a more resonant, baritone timbre (Fitch & Giedd, 1999). This sexual dimorphism in F0 and formant frequencies largely results from testosterone-related changes in size and morphology that occur during male puberty (Busby & Plant, 1995). The lengthening of the vocal folds associated with laryngeal growth causes a dramatic drop of F0 in adolescent males to one octave lower (100–120 Hz: Simpson, 2009) than females' F0 (200–220 Hz: Simpson, 2009). In parallel, the lengthening of the vocal tract that follows the overall differential body growth (Fitch & Giedd, 1999), and that is accentuated by a secondary laryngeal descent in adolescent males, results in a 1.2 ratio of female to male formant frequencies (Hillenbrand et al., 1995).

However, static, biological factors do not explain the entirety of voice gender differences. Pre-pubertal children's voices are dimorphic despite the absence of substantial differences in the morphology or dimensions of the vocal apparatus between sexes (Fitch & Giedd, 1999; Titze, 1994; Vorperian et al., 2005): boys' voices are consistently characterised by lower formants than girls' (Lee et al., 1999; Vorperian & Kent, 2007). This suggests that children acquire gender-specific articulatory behaviours that enable them to mimic the sexual dimorphism in formant frequencies present in adults. More specifically, it has been suggested that children make small adjustments to the length of their vocal tract, thereby altering formant frequency spacing and feminising or masculinising their voices (Mattingly, 1966; Sachs et al., 1973; Vorperian & Kent, 2007).

Similar vocal gestures may continue to play a role in adults, particularly where there is an explicit or implicit drive to accentuate or downplay the biologically determined voice in order to adapt to specific sexual roles and social contexts. Indeed, while the morphological dimorphism accounts for a substantial part of the voice differences between adult males and females, it cannot fully explain the intra-sexual differences in the femininity or masculinity of the voices of individuals (Rendall et al., 2008). Moreover, recent research on homosexual speech presents acoustic data that suggests the involvement of such adjustments in the expression of sexual orientation. While, contrary to popular stereotypes, 'gay speech' does not systematically reflect opposite sex patterns (Munson et al., 2006; Pierrehumbert et al., 2004), homosexual voices do display some characteristics associated with the opposite sex (e.g. sex-specific vowel formant values), even after controlling for body size (height and weight - Munson & Babel, 2007; Munson et al., 2006; Pierrehumbert et al., 2004; Rendall et al., 2008). Further, perceptual studies (Gaudio, 1994; Munson & Babel, 2007; Smyth et al., 2003) have shown that listeners' rating of the masculinity/femininity of a speaker's voice is correlated with their judgment of the speaker's sexual orientation: voices rated with higher femininity scores are more likely be judged as belonging to homosexual male speakers (and vice versa).

Here we investigate whether actors playing homosexual and heterosexual male characters in American TV modify their voice in line with gender stereotypes. More specifically, we hypothesise that actors playing homosexual characters feminise their voice by (1) increasing their mean fundamental frequency as well as its dynamic variation and (2) raising overall formant frequencies spacing. Increases in F0 and F0 variation can be achieved by raising the rate of vocal folds vibration, and its variability, and will respectively result in higher pitched and more melodious voices. Increases in formant frequency spacing reflect a shortening of the supralaryngeal vocal tract that can be achieved by spreading the lips, which effectively shortens the anterior end of the vocal tract, and also by raising the larynx, which shortens the posterior end of the vocal tract (Riordan, 1977; Titze, 1994). An increase in formant spacing results in a less resonant or baritone timbre (Fitch & Giedd, 1999).

Method

Selection of Stimuli

We identified actors who played at least one homosexual role and one heterosexual role in American television comedies or dramas, and for which at least one interview was also available. The list of suitable programs (available from the authors upon request) was compiled from Wyatt (2008), the Internet Movie Database (http://www.imdb.com/), and television networks listings.

The characters' audio samples were extracted from randomly selected episodes from Home DVDs (PAL) and TV show recordings. Interviews were selected from talk shows and DVD-extras to match the genre of the actors' selected roles. All audio samples were extracted using iSkySoft DVD Audio Ripper 1.8.2.7 (Wondershare Software Co. Ltd 2010) and saved in WAV format (sample rate 44,100 Hz, bit rate 128 kbps).

For each actor, we used approximately five audio samples from homosexual roles (*SD* = 1.4), with average duration 426s (*SD* = 210s), five audio samples from heterosexual roles (*SD* = .7) with average duration 410s (*SD* = 180s) and five audio samples from interviews (*SD* = 1.8), with average duration 531s (*SD* = 19s). The criteria for sample selection were: no background noise (e.g. music, other people speaking), no crowded settings (e.g. office, bar) and no strong emotional content. As expressive speech could not be completely avoided, samples were categorised by two listeners into five categories: emotional neutrality, fear, anger, happiness, and sadness (following Costanzo et al., 1969; Frick, 1986; Johnson et al., 1986). The few samples for which agreement was not reached (4%) were discarded. Samples were then selected in order to balance the emotional content across the two acted contexts. In the resulting dataset, homosexual and heterosexual character speech samples contained the same percentage of mild happiness (8%) and mild anger (17%) (Table 5.8.1). Samples were also selected to minimize content- dependent phonetic biases, by ensuring that vowels were similarly distributed across the three contexts. Pearson's Chi-squares showed no significant relationship between recording context and vowel type in either comedy $\chi^2(20) = 30.52$, $p > .05$ or drama, $\chi^2(20) = 27.94$, $p > .05$. The resulting data set consisted of a total of 200 samples from 15 actors, eight actors playing in comedies and seven playing in dramas. Samples were then randomly assigned to a numeric code and renamed accordingly, to ensure blind analysis.

Table 5.8.1

Distribution of emotional content between recording contexts

Emotions	Homosexual characters (%)	Heterosexual characters (%)	Interviews (%)
Happiness	8	8	6
Neutral	75	75	88
Anger	17	17	6

Acoustic Analyses

All acoustic analyses (extraction of F0 contours and formant center frequencies) were conducted with Praat 5.1.19 (Boersma & Weenink, 2009) using a custom written script (available from the authors on request). The script allows the experimenter to set all analysis parameters prior to processing, and to modify them manually if necessary (blind to sample and condition), to correct for tracking errors.

Fundamental Frequency

The script uses the built-in autocorrelation algorithm ("to Pitch" command) to extract the F0 contour, and then computes the mean (F0mean) and the standard deviation (F0SD). The analysis parameters were set as follows: pitch floor = 65Hz, pitch ceiling = 300Hz and time step = .01s. The coefficient of variation (F0CV) was then calculated, as the ratio of SD to the mean. F0CV describes the dispersion of F0 in a way that does not depend on its magnitude and thus corrects for correlative increases of F0SD with mean F0 and accounts for its logarithmic perception by human listeners (Gaudio, 1994): voices with large F0CV are perceived as more melodious than those with small F0CV (Devillers & Vasilescu, 2003; Hodges-Simeon et al., 2010).

Formants

The script uses Linear Predictive Coding (LPC: “To Formants (Burg)” command) to estimate the centre frequencies of the first four formants (F1–F4). The analysis parameters were set as follows: maximum number of formants to be extracted = 4, ceiling of the formant search range = 4,000Hz, and effective duration of the analysis window = .03s. In order to check the accuracy of formant tracking, the script displays a PRAAT Editor window (narrow band spectrogram with overlaid formant tracks) for each sample. In 12 samples, the tracks of the estimated formants were clearly not aligned with the formants visible in the spectrogram, indicating that the chosen number of poles with LPC analysis was inadequate. The ceiling of the formant search range (‘‘maximum formant’’ parameter) was raised by 200 Hz-increments to match the formant tracks with the formants on the spectrogram, from the ‘‘Formant Settings…’’ dialogue in the Editor window.

Formant Spacing

Formant spacing (ΔF) is the average interval (in Hz) between each adjacent pair of formants. It is determined by, and inversely correlated to, the length of the vocal tract of the speaker (Titze, 1994). We estimated ΔF by modelling the vocal tract as a straight uniform tube closed at the glottis and opened at the lips (full details are given in ‘‘Appendix A’’). This method of estimating ΔF is justified by the observation that, although formants vary from vowel to vowel, formant spacing (ΔF) approaches a constant determined by vocal tract length at supra-segmental level (Titze, 1994). Furthermore, psychoacoustic experiments varying ΔF have shown that the linear scaling of formant spacing determines the perceived age, size, and gender of human voice by listeners (Feinberg et al., 2005; Pisanski & Rendall, 2011; Smith & Patterson, 2005).

Statistical Analyses

For each acoustic parameter, a two-way mixed ANOVA was carried out with genre as the group factor (comedy, drama) and context (heterosexual, interview, homosexual) as the repeated factor. Post-hoc analyses were conducted by applying contrasts to study differences between the three contexts and between the two genres. All calculations and graphics were completed using SPSS v.16 for Mac.

Results

Mean Fundamental Frequency

There was a significant main effect of context, $F(2, 26) = 19.72$, $p < .001$ (see Figure 5.8.1a). Contrasts revealed that F0mean in homosexual roles ($M = 143.31$, $SD = 13.42$) was significantly higher than in heterosexual roles ($M = 122.93$, $SD = 16.85$), $F(1, 13) = 38.27$, $p < .001$, and in interviews ($M = 112.75$, $SD = 21.38$), $F(1,13) = 47.64$, $p < .001$. F0mean was not statistically different between heterosexual roles and interviews $F(1, 13) = 2.42$, $p > .05$. A main effect of genre was also found, $F(1, 13) = 5.25$, $p = .04$. However, while F0mean was significantly higher in comedy ($M = 150.99$, $SD = 12.02$) than in drama ($M = 134.52$, $SD = 9.12$), parameter estimates show that this difference was only statistically significant for homosexual roles, $t(13) = 2.95$, $p = .01$. There was no significant interaction between context and genre, $F(2, 26) = .06$, $p > .05$.

Fundamental Frequency Variation

There was a significant main effect of context on F0 standard deviation (F0SD), $F(2, 26) = 12.56$, $p < .001$ (see Figure 5.8.1b). Contrasts revealed that F0SD was significantly higher in homosexual roles ($M = 29.99$, $SD = 8.20$) than in heterosexual roles ($M = 22.33$, $SD = 7.65$), $F(1, 13) = 10.65$, $p = .006$, and interviews ($M = 18.92$, $SD = 6.64$), $F(1,13) = 25.74$, $p < .001$. F0SD was not statistically different between heterosexual roles and interviews $F(1, 13) = 2.42$, $p > .05$. There was also a main effect of genre $F(1, 13) = 6.11$, $p = .028$, with higher F0SD in comedy roles than in drama roles.

A significant main effect of context was also found on normalised F0 variation (F0CV), $F(2, 26) = 4.68$, $p = .018$ (see Figure 5.8.1c). Contrasts revealed that F0CV in actors playing homosexual roles ($M = .21$, $SD = .05$) was higher than in interviews (M = .16, $SD = .04$); $F(1, 13) = 12.79$, $p = .003$ while F0CV was not statistically different between heterosexual characters ($M = .18$, $SD = .05$) and interviews $F(1, 13) = 1.16$, $p > .05$. However, in contrast with F0SD, F0CV was not significantly different between homosexual and heterosexual roles, $F(1, 13) = 2.87$, $p > .05$. Finally, there was no significant main effect of genre on F0CV, $F(1, 13) = 2.78$, $p > .05$.

Formant Frequencies

The mean formant values (F1–F4) measured across the contexts are presented in ''Appendix B''. When playing homosexual roles, actors' voices were characterised by higher F1, F2, and F4 formants than when playing heterosexual roles (F1: $F(1, 13) = 13.038$, $p = .003$; F2: $F(1, 13) = 14.17$, $p = .002$; F3: $F(1, 13) = 4.47$, $p = .054$; F4: $F(1, 13) = 7.11$, $p = .019$) or during interviews (F1: $F(1, 13) = 41.09$, $p < .001$; F2: $F(1,13) = 10.62$, $p = .006$; F3: $F(1, 13) = 20.49$, $p < .001$; F4: $F(1, 13) = 9.64$, $p = .008$). Contrasts revealed that F3 was also higher in homosexual acted speech than in interviews $F(1,13) = 20.49$, $p = .001$, and between the two acting contexts in the comedy genre, $F(1,13) = 8.49$, $p = .012$. Furthermore, in homosexual roles, F3 and F4 were significantly higher in the comedy genre than in the drama genre (F1: $F(1, 13) = 2.22$, $p > .05$, F2: $F(1, 13) = .44$, $p > .05$, F3: $F(1, 13) = 5.89$, $p = .03$, F4: $F(1, 13) = 7.01$, $p = .02$).

Formant Spacing

There was a significant main effect of context on ΔF, $F(2.26) = 16.00$, $p = .002$ (see Figure 5.8.1d). Contrasts revealed that actors playing homosexual roles (M = 1,035.16 Hz, SD = 17.5 Hz) spoke with a higher ΔF, than when playing heterosexual roles (M = 1,009.47 Hz, SD = 24.94 Hz), $F(1, 13) = 14.98$, $p = .002$, or than when interviewed (M = 991.19 Hz, SD = 28.33 Hz), $F(1, 13) = 30.22$, $p < .001$. While actors playing heterosexual roles spoke with a slightly higher ΔF than when being interviewed, this difference approached significance $F(1, 13) = 4.61$, $p = .051$.

Furthermore, parameter estimates showed that while in homosexual roles, ΔF was significantly higher in the comedy genre than in drama, $t(13) = 2.63$, $p = .021$, in heterosexual roles, genre had no significant effect on ΔF $t(13) = .28$, $p > .05$. Finally the context by genre interaction had no significant effect on ΔF, $F(2, 26) = .56$, $p > .05$.

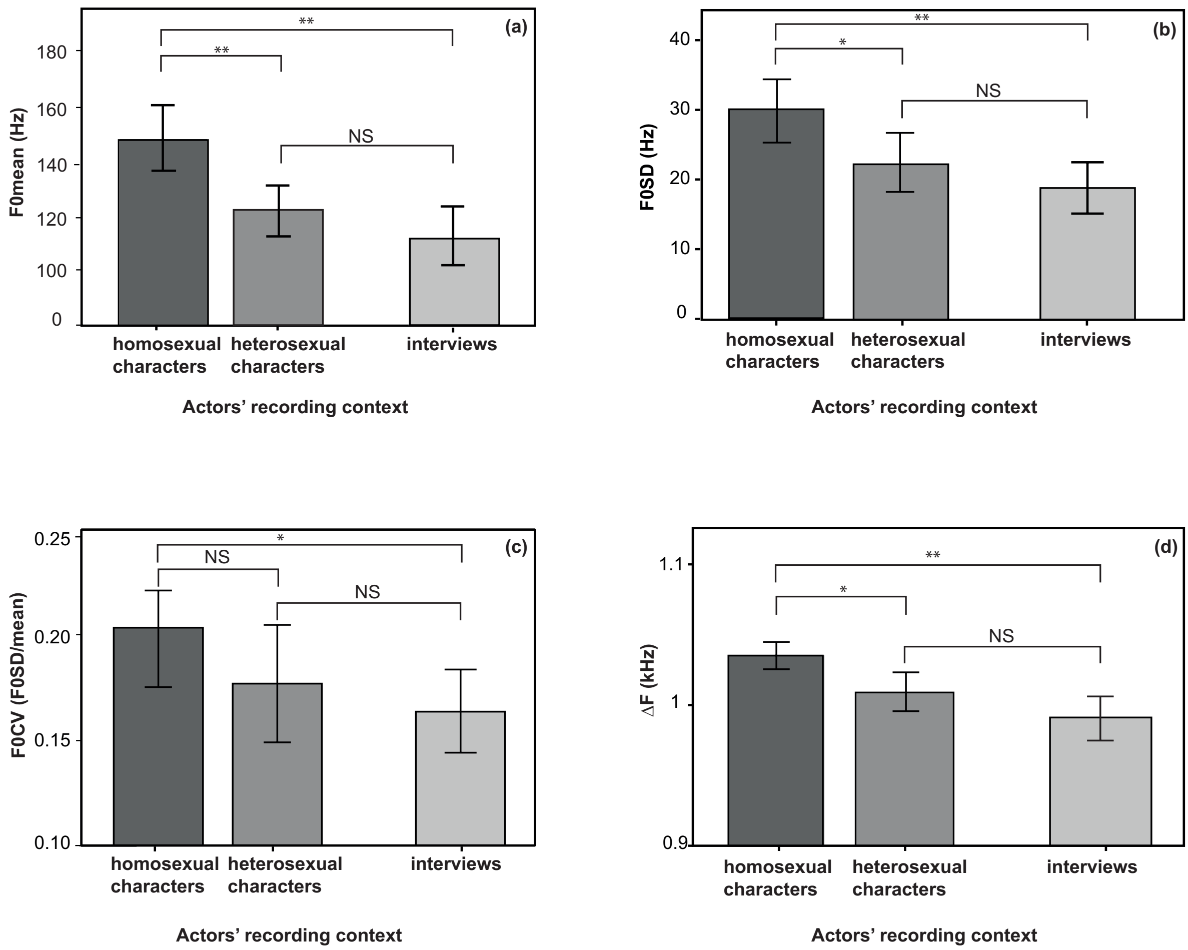

Figure 5.8.1. Mean values of (a) fundamental frequency (F0mean); (b) F0 standard deviation (F0SD); (c) coefficient of variation (F0CV) and (d) formant spacing (ΔF)— error bars: 95% CI for actors playing homosexual and heterosexual roles (across genres), and being interviewed. NS not significant, *p < .05, **p < .001

Discussion

We found that actors playing homosexual characters produced higher pitched, more melodious, and less baritone voices, by respectively increasing the mean F0, F0 variation, and formant frequency spacing of their voice. To the extent that adult female voices are characterised by higher F0 and formants than adult male voices (Titze, 1994), these manipulations created voice profiles that were less masculine and more feminine. Moreover the increased F0 variation observed in the voice of actors playing homosexual characters suggests that they attempt to increase the melodic quality of their voice, another stereotypical correlate of perceived femininity (Avery & Liss, 1994; Henton 1989, 1995; Terengo, 1966). These results confirm that the stereotypical portrayal of male homosexuals by the media, which attributes feminine values to their appearance and behaviour (Chung, 2007), also involves the feminisation of their voices. Our observation that actors playing homosexual roles in comedy further accentuated the feminine qualities of their voice is in line with previous research showing that homosexual male characters in comedy draw from stereotypes of femininity (Battles and Hilton-Morrow, 2002), as well as being the subject of jokes based on these stereotypes (Battles & Hilton-Morrow, 2002; Dow, 2001).

The frequency values achieved by actors in the different contexts can be examined in the context of published differences between men and women and between homosexual and heterosexual male speakers. While the F0 reported for heterosexual roles (123Hz) and interviews (113Hz) was comparable to that reported in male speakers (100–120Hz: Simpson, 2009), for homosexual roles (143Hz) it remained within the range of male values, but was shifted approximately 40% towards female values (200–220Hz (Simpson, 2009)). This is despite the fact that no differences in F0 are reported between homosexual and heterosexual male voices (Rendall et al., 2008). Fundamental frequency variability (speech melody) expressed as standard deviation from mean F0 (F0SD), was approximately 8Hz higher when actors played homosexual roles (30Hz) than when they played heterosexual roles (23Hz) or were being interviewed (19Hz). When F0 variability was expressed as the coefficient of variation (F0CV), homosexual acted speech remained characterised by the highest F0 variability. However, the difference was only significant when compared to interview recordings. The fact that overall, F0 variability was higher in homosexual characters than in heterosexual characters or than in interviews is consistent with stereotypical notions that women's voices are more melodious than men's (Avery & Liss, 1994; Henton, 1989; Terengo, 1966) and that less monotonous voices are perceived as more feminine (Ko et al., 2006). However this stereotype is only partly supported by acoustic studies, which do not consistently identify significant differences in F0 variation between sexes (Simpson, 2009), nor between homosexual and heterosexual male voices (Gaudio, 1994; Munson & Babel, 2007; Smyth et al., 2003) Formant spacing (ΔF) values in actors playing heterosexual roles (1,009Hz) and interviews (991 Hz) are comparable to heterosexual men's ΔF reported in the literature (1,005 Hz in Feinberg et al., 2005; and 991Hz, as calculated from F1–F4 values in Pisanski & Rendall, 2011). The ΔF observed in homosexual roles (1,035Hz) is approximately 26Hz higher than in heterosexual roles. While this ΔF represents a

14.8% shift towards normal adult female values (1,167Hz, as estimated from formant values in Pisanski & Rendall, 2011), it remains well within the normal range for adult male speakers. The observed ΔF values for homosexual acted speech are also higher than those reported in the voices of self-identified homosexual speakers (1,005Hz, as estimated from formant values in Pisanski & Rendall, 2011).

The voice stereotype identified here is likely to result from interactions between existing acoustic cues to gender and sexual orientation in non-acted speech, and perceptual and cultural biases affecting audience expectations (Hajek & Howard, 2005). Production studies on homosexual male speech have identified a partial shift of frequency-related components towards female values (Gaudio, 1994; Munson et al., 2006; Rendall et al., 2008) and voice perception studies have found that self-identified sexual orientation was a strong predictor of how listeners rate speakers' sexual orientation (Gaudio, 1994; Munson et al., 2006) and femininity (Munson & Babel, 2007; Riordan, 1977) from their voice. The likely acoustic bases for such observations can be described in terms of the source-filter theory of voice production (Fant, 1960). At the level of the source, there are no significant differences in mean F0 (Rendall et al., 2008) and F0 variability (Gaudio, 1994) between heterosexual and homosexual men. However, at the perceptual level, listeners rate male speech with higher F0 as more feminine- and gay- sounding and listeners' ratings of F0 variability correlate positively with perceived homosexuality in men (Smyth et al., 2003). While a study of vowels /IY/, /UY/, /AA/, and /AE/ embedded in four sentences spoken by Chicago-area speakers failed to find significant differences in average formant values between self-identified homosexual and heterosexual male speakers, it showed that the vocalic space was more dispersed in homosexual than heterosexual male speakers (Pierrehumbert et al., 2004). More recently, a study of homosexual males speakers from the St. Paul/Minneapolis metropolitan area found that the /AE/ and /EH/ vowels (embedded in CVC words) were characterised by higher F1 and F2 (Munson et al., 2006). Similarly, /IY/ and /UY/ were characterised by higher F1 in homosexual male speakers from southern Alberta, Canada (Rendall et al., 2008). Furthermore, perception experiments (Munson and Babel 2007; Munson et al. 2006) confirm that higher F1 and F2 values correlate with listeners' ratings of male voices as gay sounding. Thus, whilst our acoustic study suggests that acted gay speech is characterised by a shift of spectral components towards female values specific to media stereotyping, acoustic and perceptual observations of non-acted homosexual speech indicates that this shift may also partly reflect the representation and exaggeration by the media of female voice patterns adopted by some gay male speakers (Rendall et al., 2008).

Interestingly, whilst actors' voices displayed the highest frequency values for homosexual roles, the lowest values were registered for the interviews (although the difference between heterosexual characters and interviews was non-significant). Lower levels of emotional intensity and intended voice projection may account for the observed low interview values: in natural speech, vocal effort is often accompanied with an increase of fundamental frequency (Plant & Younger, 2000) and rising formant frequencies (especially F1), due to amplification of articulatory movements (Audibert et al., 2010; Tom et al., 2001). Moreover, F0 has been found to be constantly higher in acted speech than in non-acted speech, presumably due to greater levels of emotional intensity in the former (Kienast & Sendlmeier, 2000). Besides, the homosexual and heterosexual acted contexts contained higher percentages of mild emotional context ("anger" and "happiness"), which are known to raise F1 and F2 (Kienast & Sendlmeier, 2000; Murray & Arnott, 1993), suggesting that the lower values for interviews may reflect a bias due to the fewer emotional speech instances in such a context.

More generally, F0 and formant frequency adjustments similar to that identified here have been hypothesised and observed to play a role in mammal vocal communication. The "size code" theory (Ohala, 1984) posits that, across species, signallers can vary the expression of their dominance by raising F0 and formants to sound smaller, and thus less threatening, while F0 and formant lowering are associated with greater body size and aggressiveness.

Whilst studies of animal (Davis, 1987; Fitch & Reby, 2001; Lopez et al., 1988; Reby et al., 2005) and human (Puts et al., 2006) vocal communication support this hypothesis, there is also growing evidence that in human speech, F0 and formant manipulations are involved in the vocal expression of gender-related attributes (Feinberg et al. 2005; Pisanski & Rendall, 2011). As mentioned in the introduction, despite negligible differences in anatomy between the two sexes in the pre-pubertal stage (Lee et al., 1999; Sachs et al., 1973; Vorperian & Kent, 2007), boys have lower formants with consequently narrower formant spacing than girls, suggesting that children acquire the ability to behaviourally achieve gender-specific formant patterns during development (Mattingly, 1966; Sachs et al., 1973). Furthermore, increases in F0 (by shortening the length and/or increasing the tension of the vocal folds) and formant frequencies (by raising the larynx and/or spreading the lips), convey ‘friendliness’, ‘politeness’, ‘vulnerability’, and ‘femininity’ (Sachs et al., 1973), which are typically considered female characteristics, while decreases in these frequency components convey ‘aggressiveness’, ‘assertiveness’, and ‘masculinity’ (Chuenwattanapranithi et al., 2006, 2008; Puts et al., 2007).

The specific gestures at the basis of the acoustic variation reported in this study remain to be investigated. While increases in mean F0 (pitch) can be achieved by adjusting vocal fold length (Titze, 1994), upward shifts of formant frequencies (which will result in a less baritone timbre) can involve either vowel-specific or more global adjustments. For example, research on vowel fronting shows that North-American male speakers from Northern states produce raised and fronted /AE/, thus lowering F1 and raising F2 (Clopper et al., 2005), while speakers from Southern states tend to front back-vowels /UY/ and /AA/ which would lead to higher mean F1 and F2, due to the combination of tongue, lip, and laryngeal movements (Thomas, 2003). Here the upward shift is identified at supra-segmental level and involves most formant frequencies (with the exception of F3, which is not significantly raised in the drama genre), suggesting a global adjustment of vocal tract length, which could be obtained via lip spreading and/or larynx lowering. In an idealized uniform linear vocal tract with a constant cross-sectional area, vocal tract length variation by lip rounding or by larynx lowering should uniformly affect the frequency of all formants (Titze, 1994). However, vocal tract modeling (Fagel, 2010; Lasarcyk & Trouvain, 2003; Sundberg & Nordstrom, 1976) and production (Fagel, 2010; Lasarcyk & Trouvain, 2003; Tivoli & Gordon, 2008) studies show that lip and larynx movements affect formants differently and that these differences are vowel-specific. The retraction of the mouth corners (“smiling”) is characteristic of female speakers across cultures (Drahota et al., 2008; Tartter, 1980) and the effect of the associated shortening of the vocal tract on the quality of the voice has been hypothesized to contribute to the expression voice gender (Sachs et al., 1973) due to associated raising of formant frequencies (Tartter, 1980). Future studies could investigate these interactions between facial and vocal behaviours and their contribution to gender expression in general, and to the ‘‘effeminacy’’ stereotype attributed to male homosexual characters in particular.

Conclusions

This study shows that the vocal behaviour of actors playing homosexual characters conforms with the effeminacy stereotype, as they alter the frequency components of their voice along the existing sexual dimorphism in adult human voices: vocal tract resonances are raised towards female values, and F0 mean and F0 variation are increased towards female values. In perceptual terms, these manipulations result in actors having higher-pitched, lighter, and more expressive voices when playing homosexual roles than when playing heterosexual ones. These results on stereotypical acted speech show that speakers can use behavioural strategies to adjust gender-related acoustic properties at the source (F0) and filter (formants) level, in order to vary their expression of gender and gender-related attributes. The ontogeny of these vocal gestures, and the extent to which they are used for the expression of gender and sexual orientation, in both acting and everyday life, is an exciting area for future research.

Acknowledgments

Many thanks to Karen McComb and Robin Banerjee for their helpful comments on earlier versions of the paper.

References

Audibert, N., Aubergé, V., & Rilliard, A. (2010). Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: a pilot study. *In Proceedings from 5th International Conference on Speech Prosody*. Chicago, USA.

Avery, J. D., & Liss, J. M. (1994). Acoustic characteristics of less-masculine-sounding male speech. *Journal of the Acoustical Society of America, 99*, 3738–3748.

Battles, G. K., & Hilton-Morrow, W. (2002). Gay characters in conventional spaces: Will and Grace and the situation comedy. *Critical Studies in Media Communication, 19,* 87–105.

Blashill, A. J., & Powlishta, K. K. (2009). Gay stereotypes: The use of sexual orientation as a cue for gender-related attributes. *Sex Roles, 61*, 783–793.

Boersma, P., & Weenink, D. (2009). PRAAT: Doing phonetics by computer (Version 5.1.19) [Computer program]. Retrieved from: http://www.praat.org.

Busby, P., & Plant, G. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America, 97,* 2603–2606.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2006). Expressing anger and joy with the size code. In *Proceedings of Speech Prosody*. Dresden, Germany.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code: A perceptual investigation. *Phonetica, 65,* 210–230.

Chung, S. K. (2007). Media literacy art education: Deconstructing lesbian and gay stereotypes in the media. *International Journal of Art and Design Education, 26,* 98–107.

Clopper, C., Pisoni, D. B., & Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118,* 1661–1676.

Costanzo, F. S., Markel, N. N., & Costanzo, P. R. (1969). Voice quality profile and perceived emotion. *Journal of Counseling Psychology, 16,* 267–270.

Davis, M. S. (1987). Acoustically mediated neighbor recognition in the North-American bullfrog, Rana-Catesbeiana. *Behavioral Ecolology and Sociobiology, 21,* 185–190.

Devillers, L., & Vasilescu, I. (2003). Prosodic cues for emotion characterization in real-life spoken dialogs. *Proceedings of Eurospeech*, 189–192.

Dow, B. (2001). Television, and the politics of gay and lesbian visibility. *Critical Studies in Media Communication*, 18, 123–140.

Drahota, A., Costall, A., & Reddy, V. (2008). The vocal communication of different kinds of smile. *Speech Communication, 50,* 278–287.

Fagel, S. (2010). Effect of smiled speech on lips, larynx and acoustics. *Lecture Notes in Computer Science, 5967*, 294–303.

Fant, G. (1960). Acoustic theory of speech production. The Hague: Mouton and Co.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of funda- mental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69,* 561–568.

Fitch, W., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106,* 1511–1522.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *In Proceedings of the Royal Society*. London, UK: Royal Society Publishing.

Frick, R. W. (1986). The prosodic expression of anger: Differentiating threat and frustration. *Aggressive Behavior, 12,* 121–128.

Gaudio, R. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech, 96*, 30–57.

Hajek, C., & Howard, G. (2005). Intergroup communication schemas: Cognitive representations of talk with gay men. *Language & Communication, 25*, 161–181.

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication, 9,* 299–311.

Henton, C. G. (1995). Pitch dynamism in female and male speech. *Language & Communication, 15,* 43–61. Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97,* 3099–3111.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature, 21*, 406–427.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *Journal of the Acoustical Society of America, 96*, 2646–2654.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry, 43*, 280–283.

Kienast, M., & Sendlmeier, W. F. (2000). Acoustical analysis of spectral and temporal changes in emotional speech. *Proceedings from the ISCA ITRW on Speech and Emotion*. Newcastle, UK: Textflow.

Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly, 11,* 83–96.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within- and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin, 32*, 806–819.

Lasarcyk, E., & Trouvain, J. (2003). Spread lips ? raised larynx ? higher F0 = smiled speech?: An articulatory synthesis approach. In *Proceedings of the 8th International Seminar on Speech Production.* Strasbourg, France: INRIA.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America, 105,* 1455–1468.

Linneman, T. J. (2008). How do you solve a problem like Will Truman? The feminisation of gay masculinities on Will and Grace. *Men and Masculinities, 10,* 583–603.

Lopez, P. T., Narins, P. M., Lewis, E. R., & Moore, S. W. (1988). Acoustically induced call modification in the white-lipped frog, Leptodactylus albilabris. *Animal Behavior, 36,* 1295–1308.

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles, 37,* 663–685.

Mattingly, I. G. (1966). Speaker variation and vocal-tract Size (A). *The Journal of the Acoustical Society of America, 39,* 1219(A).

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass, 1,* 416–449.

Munson, B., McDonald, E., & DeBoe, N. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics, 24*, 202–240.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of Acoustic Society of America, 93*, 1097–1106.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41,* 1–16.

Perry, T., Ohde, R., & Ashmead, D. (2001). The acoustic bases for gender identification from children's voices. Journal of the Acoustical Society of America, 109, 2988–2998.

Pierrehumbert, J., Bent, T., & Munson, B. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America, 116,* 1905–1908.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *Journal of the Acoustical Society of America, 129,* 2201–2212.

Plant, R. L., & Younger, R. M. (2000). The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. *Journal of Voice, 14*, 170–177.

Puts, D., Gaulin, S., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*, 283–296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28,* 340–344.

Raley, A. B., & Lucas, L. J. (2006). Stereotype or success? Prime-time television's portrayals of gay male, lesbian, and bisexual characters. *Journal of Homosexuality, 51*, 19–38.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behavior, 65*, 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*, 944–955.

Rendall, D., Vasey, P., & McKenzie, J. (2008). The Queen's English: An alternative, biosocial hypothesis for the distinctive features of "gay speech". *Archives of Sexual Behavior, 37*, 188–204.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America, 62,* 998–1002.

Sachs, J., Lieberman, P., & Erikson, D. (1973). Anatomical and cultural determinants of male and female speech. In R. W. Shuy & R. W. Fasold (Eds.), *Language* *attitudes* (pp. 74–84). Washington: Georgetown University Press.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass, 3*, 621–640.

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex, and age. *The Journal of the Acoustical Society of America, 118*, 3177–3186.

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society, 32*, 329–350.

Staats, G. R. (1978). Stereotype content and social distance: Changing views of homosexuality. *Journal of Homosexuality, 4*, 15–27.

Sundberg, J., & Nordstrom, P. E. (1976). Raised and lowered larynx: The effect on vowel formant frequencies. *Speech Transmission Laboratory Quarterly Progress, 17*, 35–39.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics, 27*, 24–27.

Terengo, L. (1966). Pitch and duration characteristics of the oral reading of males on a masculinity- femininity dimension. *Journal of Speech and Hearing Research, 9*, 590–595.

Thomas, E. R. (2003). Secrets revealed by southern vowel shifting. *American Speech, 78*, 150–170.

Titze, I. R. (1994). Principles of voice production. Iowa City: IA7 National Center for Voice and Speech. Tivoli, M., & Gordon, M. J. (2008). The [ʔ spread] of the Northern Cities Shift. *University of Pennsylvania Working Papers in Linguistics, 14*, 111–120.

Tom, K., Titze, I. R., & Hoffman, E. A. (2001). Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness. *The Journal of Acoustical Society of America, 109*, 742–747.

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*, 1510–1545.

Vorperian, H. K., Kent, R. D., & Lindstrom, M. J. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The*

Journal of the Acoustical Society of America, 117, 338–350.

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexual. *Journal of Speech and Hearing Research, 55*, 43–50.

Wondershare Software Co. Ltd. (2010). iSkysoft Audio Ripper (Version 1.8.2.7) [Computer program]. Retrieved from: http://www.iskysoft.com.

Wyatt, D. A. (2008). Gay/lesbian/bisexual television characters. [Website] Retrieved from: http://home.cc. umanitoba.ca/*wyatt/tv-characters.html

Appendix A: Calculation of formant spacing

Formant spacing (ΔF) was calculated by fitting a model that assumes that the vocal tract is an open (lips) – closed (glottis) tube with a uniform cross-section (quarter-wave resonator) to the observed formant values (Reby and McComb, 2003). In this model individual formant frequencies are inversely related to the length of the vocal tract by the following formula:

$$(1)\quad F_i = \frac{(2i-1)c}{4VTL}$$

where *c* is the speed of sound in air (approximated as 350m/s in the vocal tract), *i* is the number of the formant (i=1,2,..) and *VTL* is the length of the vocal tract (Titze, 1994).

Since the formant frequency spacing can be expressed as the difference between any two adjacent formants, ΔF is inversely related to *VTL*:

$$(2)\quad \Delta F = F_{i+1} - F_i = \frac{c}{2VTL}$$

By replacing *VTL* in Equation (1) with ΔF estimated in Equation (2), individual formants are directly related to ΔF:

$$(3)\quad F_i = \frac{(2i-1)}{2}\Delta F$$

Thus, ΔF can be derived from Equation (3) as the slope of the linear regression of observed formant frequency values F*i* (y-axis) over the expected formant positions (2*i*-1)/2 (x-axis), and with the intercept set to 0 (Reby and McComb, 2003).

Appendix B

Table 5.8.2

Acoustic characteristics (in Hz) of actors' voices in homosexual roles

ID	F0mean	F0SD	F0CV	F1	F2	F3	F4	ΔF
1	136.45	22.38	.16	658.17	1,703.44	2,567.29	3,496.64	1,025.75
2	138.66	29.46	.21	710.72	2,069.90	2,639.85	3,479.50	1,058.96
3	169.38	30.08	.18	732.49	1,868.02	2,798.73	3,506.23	1,068.42
4	141.79	35.92	.25	590.46	1,677.16	2,620.69	3,423.58	1,016.44
5	118.84	21.35	.18	661.30	1,683.68	2,556.70	3,526.29	1,028.09
6	142.55	38.23	.27	637.36	1,791.07	2,667.11	3,496.28	1,043.33
7	138.51	15.57	.11	584.81	1,851.97	2,635.30	3,457.02	1,036.10
8	131.84	21.48	.16	668.06	2,021.73	2,571.41	3,412.64	1,035.21
9	147.08	48.37	.33	663.12	1,747.89	2,685.96	3,459.51	1,036.98
10	127.77	27.17	.21	647.93	1,885.65	2,586.81	3,360.44	1,018.14
11	178.38	34.17	.19	702.62	1,900.28	2,685.17	3,524.04	1,059.47
12	151.5	28.75	.19	627.93	1,692.68	2,615.17	3,416.65	1,016.63
13	145.58	28.12	.19	626.64	1,861.49	2,633.46	3,399.26	1,027.93
14	137.32	31.46	.23	596.65	1,798.63	2,501.30	3,414.93	1,009.61
15	164.00	37.37	.23	646.24	1,772.70	2,633.49	3,545.31	1,046.40
Mean	143.21	29.99	.21	650.30	1,821.75	2,626.57	3,461.23	1,035.16
SD	13.42	8.20	.05	42.93	118.44	69.42	54.78	17.15

Note. F0mean (Hz), F0SD (Hz), F0CV (SD/mean), mean F1–F4 formant frequency values (Hz) and spacing ΔF (Hz) for each actor (ID) and across actors playing homosexual roles

Table 5.8.3

Acoustic characteristics (in Hz) of actors' voices in heterosexual roles

ID	F0mean	F0SD	F0CV	F1	F2	F3	F4	ΔF
1	120.73	18.31	.15	573.13	1,546.60	2,477.16	3,467.66	996.96
2	117.88	23.69	.20	601.72	1,683.77	2,614.79	3,400.04	1,012.55
3	127.45	17.16	.13	574.90	1,757.37	2,471.06	3,421.19	1,003.59
4	124.51	26.48	.21	618.23	1,765.44	2,598.68	3,301.90	1,000.51
5	92.37	6.11	.07	623.88	1,644.86	2,731.25	3,430.20	1,029.19
6	118.19	21.17	.18	637.22	1,763.14	2,582.43	3,547.54	1,039.80
7	114.70	13.91	.12	604.32	1,867.71	2,669.23	3,469.71	1,043.85
8	117.93	30.29	.26	615.13	1,493.35	2,606.53	3,367.10	992.80
9	117.39	27.50	.23	588.28	1,642.46	2,407.20	3,275.31	963.78
10	103.79	21.24	.20	638.19	1,713.14	2,533.68	3,245.82	980.16
11	160.04	31.80	.20	638.29	1,925.41	2,582.14	3,523.35	1,047.35
12	146.07	33.08	.23	641.70	1,726.94	2,633.55	3,429.19	1,023.68
13	144.35	28.40	.20	600.48	1,714.24	2,659.09	3,426.67	1,024.41
14	112.99	12.91	.11	555.28	1,524.64	2,455.22	3,400.85	981.22
15	125.62	23.01	.18	585.36	1,541.22	2,458.04	3,513.46	1,002.23
Mean	122.93	22.33	.18	606.41	1,687.35	2,565.38	3,414.67	1,009.47
SD	16.85	7.65	.05	27.11	124.79	96.08	88.24	24.94

Note. F0mean (Hz), F0SD (Hz), F0CV (SD/mean), mean F1–F4 formant frequency values (Hz) and spacing ΔF (Hz) for each actor (ID) and across actors playing heterosexual roles

Table 5.8.4

Acoustic characteristics (in Hz) of actors' voices in interviews

ID	F0mean	F0SD	F0CV	F1	F2	F3	F4	ΔF
1	107.73	11.83	.11	543.24	1,501.34	2,430.73	3,240.24	949.58
2	108.52	21.75	.20	589.44	1,576.22	2,264.63	3,297.41	945.79
3	156.62	27.63	.18	579.77	1,715.74	2,602.84	3,372.69	1,008.33
4	104.76	15.73	.15	563.30	1,608.11	2,507.31	3,305.99	977.77
5	87.79	13.11	.15	571.58	1,731.28	2,523.78	3,251.82	979.69
6	113.97	24.63	.22	609.04	1,560.51	2,491.88	3,555.07	1,015.13
7	100.82	11.64	.12	538.81	1,659.24	2,609.09	3,472.88	1,020.77
8	139.77	27.16	.19	619.83	1,471.86	2,527.42	3,290.53	969.20
9	100.93	21.48	.21	578.89	1,921.19	2,573.93	3,475.56	1,036.69
10	96.50	11.33	.12	536.56	1,701.42	2,460.72	3,412.12	995.94
11	98.60	15.67	.16	605.08	1,646.72	2,424.97	3,604.06	1,021.39
12	109.36	20.76	.19	605.79	1,748.47	2,708.65	3,337.66	1,018.05
13	113.64	14.20	.12	589.46	1,753.40	2,566.68	3,234.25	983.88
14	95.30	14.98	.16	545.65	1,621.10	2,411.03	3,251.98	957.81
15	156.94	31.98	.20	543.52	1,519.56	2,462.29	3,438.97	987.77
Mean	112.75	18.92	.16	574.66	1,649.08	2,504.40	3,369.42	991.19
SD	21.38	6.64	.04	28.38	117.86	104.81	118.68	28.33

Note. F0mean (Hz), F0SD (Hz), F0CV (SD/mean), mean F1–F4 formant frequency spacing ΔF (Hz) for each actor (ID) and across actors in interviews

Chapter 6: Summary of Results

This research set out to investigate speakers' ability to control the expression of their gender and related attributes through their voice, within the source-filter theory framework of vocal communication, and by extending John Ohala's (1984) "size code" theory to the vocal expression of gender ("gender code"). More specifically, I made the hypothesis that acoustic cues to gender in the human voice are not solely determined by sex and individual differences in the anatomy of the vocal apparatus, but are also affected by acquired vocal gestures enabling the conventionalised use of biologically based vocal cues to sex (F0 and ΔF). Using a variety of analysis techniques and experimental paradigms including speech analysis, speech resynthesis, psychoacoustic experiments and video analysis of facial gestures, I provided some clear evidence for this "gender code" by firstly exploring the biological correlates to acoustic and perceptual variation in voice gender (Chapter 3), then by focusing on individuals' ability to control sexually dimorphic voice cues to alter their gender expression and listeners' perceptions of such adjustments (Chapter 4), and finally by looking at the interplay between voice gender and social context, and in particular cultural stereotypes to gender and sexual orientation (Chapter 5). The following section integrates the results from my studies in the context of the six research questions underpinning this thesis, whilst highlighting main unresolved issues and suggestions for future directions.

Question 1. How does the natural variation in the gender-related acoustic cues relate to speakers' anatomical and biological differences?

Question 2. What is the perceptual relevance of this variation in terms of listeners' gendered attributions of speakers?

Both questions were explored in Chapter 3. In the first instance, to exemplify the mismatch between acoustic and biological correlates to gender expression (Question 1), Study 1 compared sex-linked developmental changes in anatomical (resting) vocal tract length (MRI-VTL) with changes in its resonant frequencies (formants). I decided to focus on formant values based on availability of data on anatomical vocal tract measurements, compared to the lack of anatomical data on vocal fold length that can be linked to its acoustic correlate, F0 (Fitch & Giedd, 1999; Vorperian et al, 2009; Vorperian et al., 2011). Moreover, unlike F0, differences between males' and females' formant values emerge well before the pubertal dimorphism in the vocal apparatus, suggesting gender differences in vocal behaviour (Lee, Potamianos & Narayanan, 1999; Busby & Plant, 1995; Perr, Ohde & Ashmead, 2001). By interpreting sex differences in individual formant values (Fi) in terms of global vocal tract length adjustments (apparent Vocal Tract Length – aVTL), I was able to demonstrate that males speak with a longer vocal tract (thus lowering their formants and narrowing their formant spacing) than females from early childhood, confirming that biologically based sex differences in vocal tract length are not sufficient to explain the observed acoustic dimorphism between males and females. Moreover, my results pointed at gender-specific behaviours in vocal tract length manipulations, such as young boys masculinising their voices and/or girls feminising theirs, thus imitating the adult sex dimorphism in this parameter.

Another pre-requisite of the "gender code" is that the naturalistic variation in voice gender cues has a functional relevance: it is attended to and used by listeners to characterise the gender of unseen speakers (Question 2). Psychoacoustic studies have already provided evidence of the perceptual relevance of sexually dimorphic cues in adult voices: listeners are influenced by small variations in F0 and ΔF when characterising speakers' gender, masculinity and femininity (Mullenix, Johnson, Topcu-Durgun & Farnsworth, 1995; Pisanski & Rendall, 2011). Similarly, Study 2 explored whether naturalistic variation in ΔF, which cues for sex in pre-pubertal voices (Perry et al., 2001), also elicited gendered perceptions of child speakers by adult listeners. The results of the sex identification and gender rating experiments showed, for the first time, that small, sex-related acoustic variation in ΔF (resynthesised within children's natural range) proportionally affects listeners' gendered attributions, with lower ΔF being consistently rated as belonging to more masculine children. It also showed that stimuli from boy exemplars were perceived as more masculine than those from girl exemplars, despite the two resynthesis continua largely overlapping, and having the same F0 and intonation, thus indicating that other acoustic cues may be at play.

Having provided some evidence that ΔF variation has a behavioural component, especially in pre-pubertal children, and shown its effect on the perceived sex, masculinity and femininity of child speakers, I then sought to explore similar relationships in adults' voices. Study 3 used a novel approach (path analysis) to simultaneously explore the links between physiological, acoustic and perceptual dimensions of men's masculinity, looking at how natural variation in sexually dimorphic voice cues (F0 and ΔF) of male speakers mediates the effects of their fitness-related characteristics (testosterone and height) on masculinity ratings made by women listeners. In testing the inter-dependence of these three dimensions, I was able to replicate previous work showing that male speakers who were taller and had higher salivary testosterone levels, also had lower F0 and ΔF (Evans, Neave, Wakelin, & Hamilton, 2008; Rendall, Kollias, Ney, & Lloyd, 2005; Puts, Apicella, & Cárdenas, 2012), and in turn, were rated as more masculine (Feinberg, 2008; Pisanski & Rendall, 2011; Pisanski et al., 2012). The study also showed that the observed inter-individual hormonal and anatomical differences did not account fully for the acoustic variation in F0 and ΔF, suggesting that some of that variation may be behavioural. It also showed that variation in F0 and ΔF did not fully account for masculinity ratings, suggesting that vocal masculinity may be expressed by other voice parameters beyond those observed.

Questions 1 and 2: Future Directions

Studies 1–3 showed that anatomical and biological factors could not fully explain acoustic gender-related variation of sexually dimorphic cues in children and adults (Question 1). They also showed that acoustic variation in the sexually dimorphic voice cues was perceptually relevant when making gendered inferences of child and adult speakers (Question 2).

At a production level, future research into the biological factors responsible for gender-related acoustic variation throughout development is warranted. For example, the mismatch between apparent and anatomical VTL (Study 1), together with reports of sex differences in vocal tract growth trend, type, and rate for select vocal tract structures at localised age ranges, call for more detailed and sophisticated techniques (e.g. 3D cine-MRI to measure dynamic changes to vocal tract dimensions during phonation) to better quantify the relationship between anatomical vocal tract morphology and its acoustic correlates during the course of development (e.g. on a year-by-year basis). Study 3 also highlighted that body height and testosterone cannot fully account for intra-individual variation in F0 and ΔF, raising the possibility that these acoustic cues may also signal for other biological indexes of masculinity. The relationship between other sexually dimorphic biological traits (e.g. facial width-to-height ratio: Lefevre et al., 2013) and acoustic variation warrants further research, particularly in women, who remain a critically understudied population (Abitbol et al., 1999). Finally, because gender has a social as well as biological dimension (Udry, 1994), social measurements should also be considered (e.g. how speakers' self-ratings of masculinity relate to their vocal expression of masculinity and to the impression they make on listeners).

At a perceptual level, while Studies 2 and 3 revealed that ΔF (in pre-pubertal children and adults) and F0 (in adults) are two key voice cues to gender attributions of speakers, they also showed that these parameters could not account fully for listeners' ratings. While my research was conducted within the source-filter and "size code" theories, and thus deliberately focused mainly on F0 and ΔF, other acoustic traits may also cue for gender. For example, women are perceived to speak with less monotonous and more breathy voices than men and speakers of both genders displaying greater intonation and breathiness in their voices are also rated as more feminine (Klatt&Klatt, 1990; Van Borsel, Vandaele, & Corthals, 2009; Wolfe, Ratusnik, Smith & Northrop, 1990), while men are reported to sound more tense (Pittam, 1987; Van Rie & Bezooijen, 1995), and creaky (Price, 1989). Including a broader spectrum of acoustic features would provide valuable insights into how gender expression is encoded in the human voice and in turn affects listeners' gendered attributions of speakers.

Question 3. Can individuals control fundamental and formant frequencies in order to vary the expression of gender, masculinity and femininity of their voice, and does the acoustic co-variation of these parameters occur along the existent sex dimorphism?

Having confirmed that anatomical and biological factors cannot fully explain acoustic gender-related variation in F0 and ΔF of children and adults, and having shown that such variation is perceptually relevant when making gendered inferences of speakers, the studies in Chapter 4 tested the specific hypothesis that speakers control the perceived gender of their voice by spontaneously varying their F0 and ΔF in line with the sex dimorphism observed in those parameters.

In Study 4, I asked pre-pubertal speakers to sound like a “boy” or a “girl” as much as possible and tested whether children increased their ΔF, which is sexually dimorphic prior to puberty, to sound more like a girl and decrease it to sound more like a boy. In line with this hypothesis, I found that both boys and girls exaggerated behavioural differences in ΔF that exist in their age group (while also confirming that boys speak with lower ΔF than girls, but had the same F0). I also found that boys raised their F0 to feminise their voices, while girls lowered F0 to masculinise theirs, despite this parameter not being sexually dimorphic prior to puberty. By extending the imitation paradigm to adult speakers (who unlike children already have a biologically dimorphic voice), Study 5 investigated whether adults would lower their sexually dimorphic cues, F0 and ΔF, to sound more “masculine” and raise them to sound more “feminine”. As predicted, I found that both men and women were capable of making those adjustments.

Figure 6.1 (child speakers) and Figure 6.2 (adult speakers) illustrate this behavioural capability in relation to speakers’ ΔF (expressed as dynamic adjustments of vocal tract length – aVTL) when speaking across conditions. Estimates of normally speaking aVTLs from a comprehensive longitudinal acoustic study (Lee et al., 1999) are added for reference. The figures show that ΔF measurements from normal speaking voices are in line with Lee and colleagues’ data, though child speakers in my study exhibited marginally longer aVTLs than Lee’s. Differences in size, sex and age distribution between the two samples may account for this discrepancy.

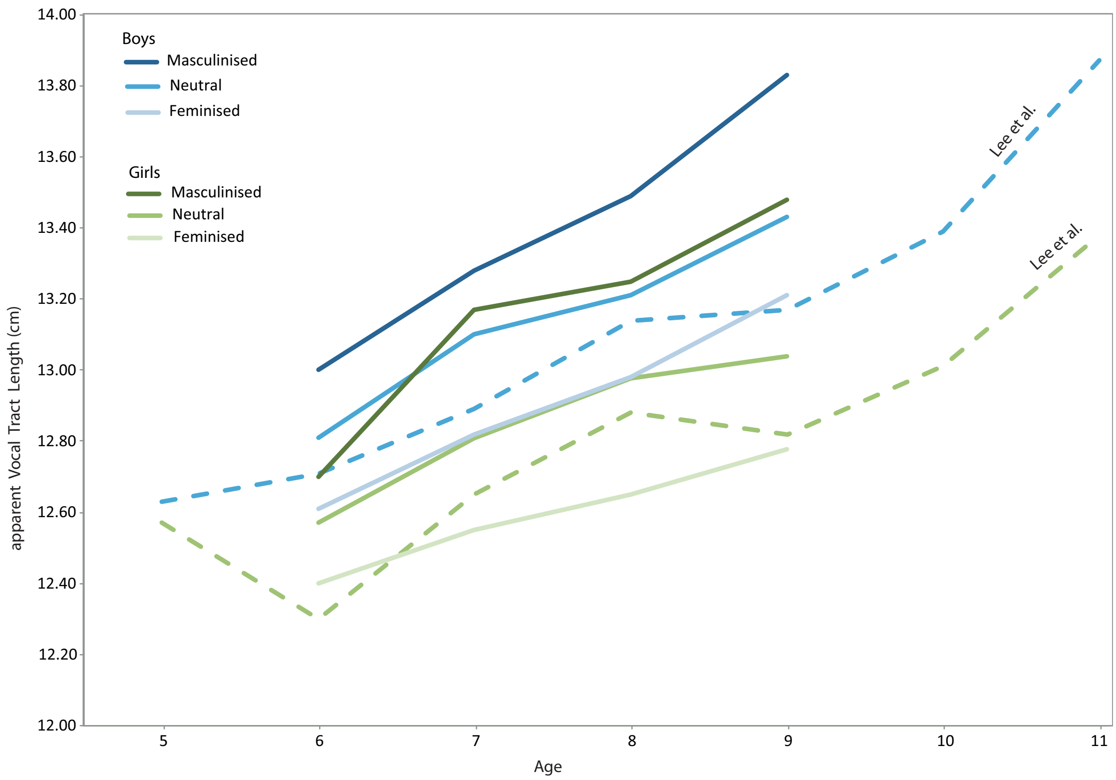

Figure 6.1. apparent Vocal Tract Length (aVTL, in cm) of pre-pubertal child speakers across vowels within each condition (masculinised, neutral, feminised). Compared to their normal speaking voices, both boys and girls significantly lengthened their aVTL (thus lowering their ΔF) when masculinising their voices (by 2% and 4% respectively) and shortened their aVTL (thus raising their ΔF) when feminising them (by 3%).

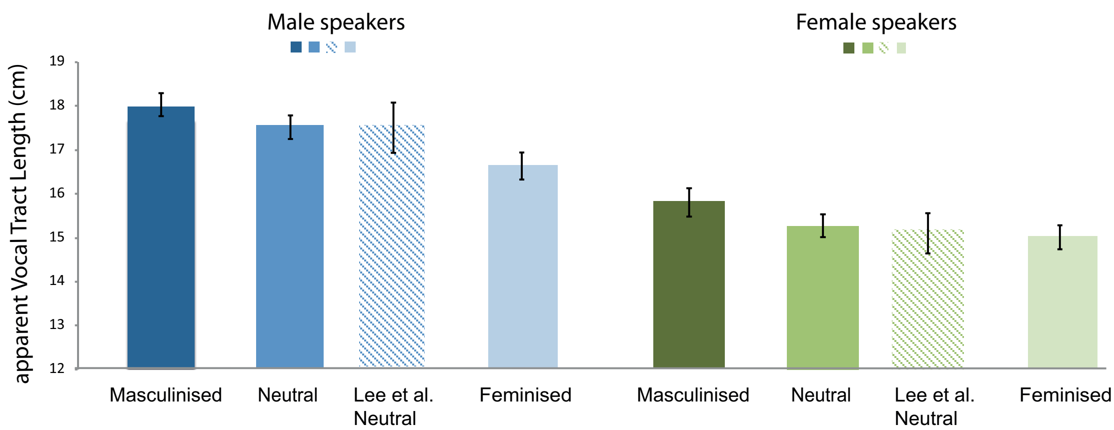

Figure 6.2. apparent Vocal Tract Length (aVTL, in cm) of adult male and female speakers across vowels within each condition (masculinised, neutral, feminised). Compared to their normal speaking voices, men significantly lengthened their aVTL (thus lowering their ΔF) by 3% when masculinising their voices, and significantly shortened it (thus raising their ΔF) by 5% when feminising them. Women significantly

lengthened their aVTL (thus lowering their ΔF) by 5% when masculinising their voices, and shortened it (thus raising their ΔF) by 1% when feminising them (albeit not significantly).

Question 4. Are speakers aware of what voice and articulatory gestures they use to vary the gender expression in their voice?

I also asked children (Study 4.1) and adult speakers (Study 5) to describe what they did to masculinise or feminise their voices. I found that adults and, though to a lesser extent, children, were aware of the perceptual outcome of their manipulations: e.g. they described making their voice sound “lower” (or “higher”) to masculinise (or feminise) it. Adults were also aware of pitch changes, and to a lesser extent, vocal tract adjustments via lip rounding/spreading (both genders) and lowering of the larynx (males only). In contrast, generally children did not report being aware of pitch and vocal tract adjustments. The observed differential awareness of vocal behaviours may reflect increased knowledge in voice differences with age, as well as developmental sex-specific processes in terms of gender identity and stereotyping (Berk, 2000; Miller, Lurye, Zosuls & Ruble 2009).

Question 5. What is the perceptual relevance of these gestures?

Study 6 investigated whether the frequency shifts reported in Studies 4 and 5 were perceptually relevant by asking listeners of both genders to make gendered attributions of children and adult speakers' vowel utterances produced in the three conditions (normal, masculinised and feminised). Consistent with my hypothesis, listeners rated lower-pitched, more resonant voices as belonging to more masculine speakers than higher-pitched, less resonant voices. In addition, listeners perceived both boys and girls as boys when masculinising their voices and girls when feminising them, revealing that children’s ability to control their gender can overcome potential anatomical differences and thus lending further support to the hypothesis that children’s acoustic dimorphism may have a behavioural origin (Lee et al., 1999; Sachs, Lieberman, & Erickson, 1973).

Questions 3 to 5: Future Directions

Taken together these results reveal that since at least six years of age, speakers can control sexually dimorphic acoustic cues in line with the existing, biologically determined dimorphism to vary the expression of their gender and related attributes through the voice. As such, these observations support the idea that gender differences in speakers' voices (especially in young children, due to the absence of anatomical dimorphism) have a behavioural dimension.

However, when and how speakers learn gender-related voice behaviours remain to be investigated. The imitation paradigm could be usefully replicated with speakers of different ages to shed further light on the age and sex-specific development of these abilities, and speakers' awareness of them. Moreover, the specific gestures at the basis of the observed acoustic variation remain to be investigated. In Study 5, objective measures of lip spreading and openness in adult speakers (taken from still images captured during the audiovisual recordings) revealed that women spoke with greater lip spreading than men, and both genders reduced their lip spreading when masculinising their voices. More sophisticated measurements of lip rounding/spreading (e.g.. via motion tracking – Yehia et al., 1998) and laryngeal lowering/raising (e.g. via 3D cine-MRI) are now needed to clarify the relationship between vocal tract adjustments and shifts in ΔF in both children and adults. Moreover, while MRI imaging has been mostly used in static imaging of the vocal tract, the same technology could be used to infer extra-laryngeal F0 control. From analysis of successive image sequences from MRI of the larynx during phonation, Honda and colleagues (1999) reported that, while source and filter components are largely independent, in line with the source-filter theory (Fant, 1960), some interplay between the two components can occur: vertically lowering the larynx simultaneously rotates the cricoid cartilage along the cervical lordosis, increasing their mass per unit and decreasing their tension. This effect overcomes the associated shortening of the vocal folds, thus lowering F0 (Figure 6.3). Changes in lung pressure and vertical head positions have also been found as extra-laryngeal mechanisms to control F0 (e.g. increasing lung pressure raises F0 by increasing tissue stress, while lowering one's head causes a rotation of the cricoid cartilage as previously described, thus lowering F0) and therefore could also be monitored (Titze, 1995; Honda, 1999; Sundberg, 1977).

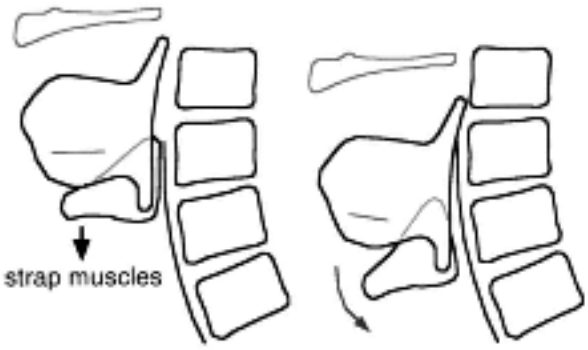

Figure 6.3. Extra-laryngeal F0 control by larynx lowering. The vertical lowering of the larynx causes a rotation of the cricoid cartilage along the cervical lordosis, resulting in shorter, but also less tense vocal folds. The decrease in tensions overcomes the associated shortening, thus lowering F0. Reproduced from “Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control”, by Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). *Language and Speech*, *42*(4), 401–411. Copyright (1999) by Sage Publications.

At the perceptual level, Study 6 also showed that listeners attend to F0 and ΔF adjustments when making gendered attributions of speakers. Future work is needed to establish the relative role of F0 and ΔF in influencing listeners’ perceptions. Previous research has found that F0 is the most salient cue in gender identification, reflecting the greater sex dimorphism in this parameter (Lass, Hughes, Bowyer, Waters & Bourne, 1976; Whiteside, 1998; Gelfer & Mikos, 2005; Hillebrand & Clark, 2009; but see Smith and Patterson, 2005). However ΔF has been found to have greater salience than F0 in within-gender attributions, when the two parameters are resynthesised by the same discernible amount (Pisanski & Rendall, 2011). Independently shifting F0 and ΔF according to the magnitude of the observed shifts would help shed light onto whether listeners weigh F0 more than ΔF or viceversa when speakers spontaneously shift such cues to masculinise or feminise their voices.

Question 6. How does the “gender code” interact with cultural stereotypes in the expression and perception of gender attributes and sexual orientation?

In line with the “gender code” hypothesis, the previous chapter found that speakers spontaneously masculinise or feminise their voices by making a conventionalised use of the voice sexual dimorphism, and that such vocal gestures are perceptually relevant to listeners. Studies in Chapter 5 aimed at investigating whether the use of the "gender code" may vary according to context, by specifically looking at its interaction with cultural stereotypes in the expression and perception of gender attributes and sexual orientation.

Psychoacoustic studies have shown that adult listeners tend to make inferences about the gender traits of adult speakers by overgeneralising sex-signalling cues in adult voices (F0, ΔF): for example, lower-pitched, lower-resonance voices are consistently attributed to more masculine individuals (Pisanski et al., 2012; Klofstad, Anderson & Peters, 2012; Anderson & Klofstad, 2012). However inferences about gender are not normally based on the evaluation of isolated cues (e.g. by listening to the voice only), but are often affected by the interaction of multiple dimensions (e.g. situational context, audio and visual sensory information). Indeed, research using visual material has consistently shown that both children and adults make stereotypic inferences by integrating information about one's gender with gendered cues in multiple domains such as choice of interests, peers, activities and appearance (Ashmore & DelBoca, 1979; Martin, Wood & Little, 1990; Banerjee & Lintern, 2000; Biernat, 1991). In line with these observations, Study 7 investigated whether children and adult listeners spontaneously linked resonance (ΔF) variation in children's voices to their gender stereotypical characterisations (e.g. choice of activities, friends and toys). Besides showing that children are sensitive to ΔF cues when making judgments of gender-related attributes in their peers (complementing what I had previously found for adult listeners in Study 2), Study 7 revealed that, overall, listeners attributed the congruent voice to the gendered characterisations: children's masculine voices were associated with stereotypically masculine portrayals (and feminine voices with stereotypically feminine portrayals). I also found that boys preferentially chose feminised voices for girls across scenarios, in line with sex-specific conceptualisations of gender (e.g. boys generally hold stronger stereotypes of girls than vice-versa (Miller et al., 2009; Berk, 2000)).

Having shown that listeners integrate voice variation and cultural stereotypes when making gendered attributions of (child) speakers, I investigated whether speakers, in turn, may also use the "gender code" to modify their voice in order to convey stereotypical portrayals to an audience. In line with the observation that homosexual male characters are stereotypically characterised by the media as having feminine mannerisms and lifestyles (Battles & Hilton-Morrow, 2002), Study 8 looked at whether this “effeminacy” stereotype was also conveyed through the voice of homosexual characters. As expected, actors playing these roles were found to raise F0 and ΔF towards female values (resulting in voices that have higher pitch and resonance). My results provide first evidence that in certain contexts, such as acting, speakers indeed use the “gender code” in order to convey cultural stereotypes (in this case gender-typed notions of sexual orientation).

Question 6: Future Directions

Study 7 provided some evidence that listeners make associations between auditory variation in the sexually dimorphic cues of the voice and gender stereotypes. However, further research is needed to clarify how vocal cues and social context inter-relate at a perceptual level. For example, it remains to be ascertained how these associations develop, and to what extent they are used when judging others in a variety of contexts (e.g. in professional vs. informal settings, in accordance with different listeners’ motivational and emotional states, and in stranger vs. familiar interactions). Additionally, further work is needed to understand the relative role of vocal cues to gender and cues expressed in other domains (e.g. what is the relative contribution of vocal and visual cues when making inferences about speakers?).

Study 8 suggested that individuals are not only aware of associations between voice variation and gender stereotypes, but they actively change their voices cues in order to project an identity that is more or less compliant with gender roles and expectations. While the present study focused on acting, future studies should explore the effect of social norms on the behavioural expression of voice gender in everyday life. For example, there is long-standing evidence that children and adults respond negatively to peers who violate traditional gender roles (Fagot, 1977; Lee & Troop-Gordon, 2011; O’Leary & Donoghue, 1978; Steinfeldt & Steinfeldt, 2012), and that as such individuals feel pressurised to endorse and adhere to them in a variety of ways, including choice of activities (Benarjee & Lintern, 2000; Buccheri, Gürber & Brühwiler, 2011), emotional displays (Ragins & Winkel, 2011) and visual appearance (Thompson, 2012). The same social pressures may also extend to the auditory domain: speakers may make differential use of vocal gestures in order to accentuate or minimise the expression of their voice gender according to the strength and characterisation of gender roles specific to the society they live in, and in the presence of peers or adult models. Finally, future work is needed to shed light on the extent to which vocal gestures used to express gender may contribute to the nonverbal maintenance of stereotypical characterisations. For example, lip spreading to feminise one's voice should result in more smiling facial expressions (Ohala, 1984) and indeed females are reported to speak with a smile in several cultures (Drahota, Costal & Reddy, 2008; Hecht & LaFrance, 1998). At the same time, individuals who smile are perceived as more feminine, warmer, less assertive and less powerful (Deutch, 1990; Deutch, LeBaron, & Fryer, 1987; Frieze & Ramsey, 1976; Kawamura & Kageyama, 2006).

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, *13*(3), 424–446.

Anderson, R. C., & Klofstad, C. A. (2012). Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles. *PLoS ONE*, *7*(12), e51216. doi:10.1371/journal.pone.0051216

Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—Social psychological conceptualization. *Sex Roles*, *5*(2), 219–248.

Battles, K., & Hilton-Morrow, W. (2002). Gay characters in conventional spaces: Will and Grace and the situation comedy genre. *Critical Studies in Media Communication*, *19*(1), 87–105.

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408.

Berk, L. E. (2000). *Child Development : 5th Edition*. Allyn and Bacon.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: a developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351–365.

Buccheri, G., Gürber, N. A., & Brühwiler, C. (2011). The impact of gender on interest in science topics and the choice of scientific and technical vocations. *International Journal of Science Education*, *33*(1), 159–178.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*, *97*(4), 2603–2606. doi:10.1121/1.412975

Deutsch, F. M. (1990). Status, Sex, and Smiling The Effect of Role on Smiling in Men and Women. *Personality and Social Psychology Bulletin*, *16*(3), 531–540.

Deutsch, F. M., LeBaron, D., & Fryer, M. M. (1987). What is in a Smile? *Psychology of Women Quarterly*, *11*(3), 341–352. doi:10.1111/j.1471-6402.1987.tb00908.x

Drahota, A., Costall, A. & Reddy, V. (2008). The vocal communication of different kinds of smile. *Speech Communication*, *50* (4), 278–287.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4), 783–788.

Fagot, B. I. (1977). Teachers' reinforcement of sex-preferred behaviours in Dutch Preschools. *Psychological Reports*, *41*(3f), 1249–1250.

Fant G (1960) Acoustic Theory of Speech Production. The Hague: Mouton & Co.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(2), 112–118.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511. doi:10.1121/1.427148

Frieze, I. H., & Ramsey, S. J. (1976). Nonverbal Maintenance of Traditional Sex Roles. *Journal of Social Issues*, *32*(3), 133–141.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, *19*(4), 544–554.

Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, *24*, 1332–1342.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control. *Language and Speech*, *42*(4), 401–411. doi:10.1177/00238309990420040301

Kawamura, S., & Kageyama, K. (2006). Smiling Faces Rated More Feminine Than

Serious Faces in Japan. *Perceptual and Motor Skills*, *103*(1), 210–214.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820–857.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704. doi:10.1098/rspb.2012.0311

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, *59*(3), 675–678.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455.

Lee, E. A. E., & Troop-Gordon, W. (2011). Peer processes and gender role development: Changes in gender atypicality related to negative peer treatment and children's friendships. *Sex Roles*, *64*(1-2), 90–102.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, *34*(4), 273–279.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development*, *61*(6), 1891–1904.

Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of Gender Stereotype Domains: Developmental and Gender Differences in Children. *Sex Roles*, *60*(11-12), 870–881.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, *98*(6), 3080–3095.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. doi:10.1159/000261706

O'Leary, V. E., & Donoghue, J. M. (1978). Latitudes of Masculinity: Reactions to Sex-Role Deviance in Men. *Journal of Social Issues*, *34*(1), 17–28.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*, 2988.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–2212.

Pittam, J. (1987). The long-term spectral measurement of voice quality as a social and personality marker: a review. *Language and Speech*, *30*(1), 1–12.

Price, P. J. (1989). Male and female voice source characteristics: Inverse filtering results. *Speech Communication*, *8*(3), 261–277.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 601–609.

Ragins, B. R., & Winkel, D. E. (2011). Gender, emotion and power in work relationships. *Human Resource Management Review*, *21*(4), 377–393.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*, *117*(2), 944–955.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age a). *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186.

Steinfeldt, J. A., & Steinfeldt, M. C. (2010). Gender role conflict, athletic identity, and help-seeking among high school football players. *Journal of Applied Sport Psychology*, *22*(3), 262–273.

Sundberg, J. (1977). *The acoustics of the singing voice*. Scientific American.

Thompson, L. J. (2012). *Becoming what women want: formations of masculinity in* *postfeminist film and television*. University of Warwick.

Udry, J. R. (1994). The nature of gender. *Demography*, *31*(4), 561–573.

Van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and Pitch Variation in Lesbian Women. *Journal of Voice*, *27*(5), 656.e13–656.e16.

Van Rie, J., & Van Bezooijen, R. (1995). Perceptual characteristics of voice quality in Dutch males and females from 9 to 85 years. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 290–293).

Whiteside, S. P. (1998). The identification of a speaker's sex from synthesized vowels. *Perceptual and Motor Skills*, *87*(2), 595–600.

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders*, *55*(1), 43–50.

Chapter 7: General Discussion

The human voice is sexually dimorphic with adult males speaking with a considerably lower fundamental frequency (F0 – the primary acoustic correlate of pitch), and lower, more closely spaced resonances or formants (ΔF – affecting perceptions of timbre) than adult females, and pre-pubertal boys speaking with lower-resonance voices than girls (Titze, 1994). As mentioned at the start of my thesis, the anatomical origins of sex voice differences are well understood, at least for adults: in addition to the body size dimorphism (men are bigger than women on average (Gaulin & Boster, 1985)), pubertal males develop a disproportionally larger larynx, which lengthens the vocal folds, thus producing lower F0, and longer vocal tract, thus producing lower, more closely spaced formants (Titze, 1994). At the same time, biological factors cannot solely account for the observed sex-related acoustic variation. For example, it has been observed that adult differences in F0 and ΔF exceed differences in size within and between sexes (Johnson, 2006). Moreover, the anatomical origin for formant differences between pre-pubertal boys and girls remains largely unknown (Vorperian et al., 2011).

Throughout my thesis, I argue and provide evidence for a behavioural role in sex-related acoustic variation by showing that on top of static, biologically determined sex differences in the voice, individuals can use a "gender code", dynamically adjusting the sexually dimorphic cues of their voices in order to deemphasise or accentuate the apparent gender of their voice and related attributes (our femininity, our masculinity). By applying the source-filter theory (Fant, 1960) to the analysis of voice gender expression, this hypothesis links F0 and ΔF dimorphisms to underlying differences in their production mechanisms, thereby providing a sound methodological framework to systematically investigate the contribution of biological and behavioural factors to voice gender variation and its ultimate effect on listeners. However, the importance of the "gender code" is not just methodological. Over the last five years, anatomical and behavioural aspects of human voice dimorphism have received increasing attention from evolutionary psychologists. Indeed, by combining acoustic and perceptual investigations with comparative studies on non-human vocal communication (Fitch, 2000; Pisanski, Mishra & Rendall, 2012), the field of Evolutionary Psychology has provided valuable insights into the functional origins of sex differences in the voice (how F0 and ΔF are both sexually selected as they advertise personal dimensions relevant to mate competition and/or choice). Yet, by interpreting sex voice differences in terms of their relevance within agonistic and sexual interactions, the evolutionary perspective fails to recognise that existing differences between and within the sexes result from on-going socialisation experiences that are time- and culture-specific (Flaherty & Richman, 1989). Therefore, a broader perspective in accounting for human voice dimorphism was needed. By transposing the "size code" hypothesis to the context of voice gender expression, the "gender code" provides a valuable theoretical framework that reconciles evolutionary explanations for sex differences in the human voice with a social understanding of such differences, and in particular how vocal behaviours may be socially enacted, produced, established and constructed to convey gender-related meanings which go beyond selective processes. In this last section I want to emphasise the importance of such unifying framework by reviewing recent contributions of evolutionary theory in understanding the evolutionary origins of voice gender differences, as well as highlighting the limitations of looking at voice gender exclusively through evolutionary lenses.

Variation in sexually dimorphic voice cues: an evolutionary prospective

Evolutionary theory (Darwin, 1871) states that certain sex differences (e.g. size, shape, colour) may have evolved through sexual selection, the evolutionary process by which individuals try to successfully select a mate in order to reproduce. In line with this theory, a growing body of research has sought to explain the existing sex dimorphism in the human voice in terms of ancestral adaptations to sexual selection pressures by investigating the role of acoustic signals in courtship and competitive behaviours in human and non-human animals (Hodges-Simeon et al., 2011). In support of this hypothesis, it has been observed that males of many polygynous species (including humans) are characterised by lower frequency vocalisations than females, due to males developing a bigger vocal apparatus than females during sexual maturity, under the influence of sex hormones and timed to influence the process of mate choice (e.g. fallow deer: Fitch & Reby, 2001; Mongolian gazelle: Frey et al., 2008; and gorillas: Dixson, 2012). Moreover, inter-individual acoustic variation appears to relate to key ecological traits of callers. For example, body size is negatively related to the fundamental frequency of call in toads, frogs and birds (Davies & Halliday, 1978; Bee et al., 1999;

Ryan & Brenowitz, 1985), while the overall spacing between formants is closely and negatively correlated to caller size in mammals (humans: Evans et al., 2006; domestic dogs: Riede & Fitch, 1999; red deer: Fitch & Reby, 2001; Reby & McComb, 2003; rhesus macaques: Fitch, 1997; giant pandas: Charlton et al., 2009). In mammals, F0 is also negatively and closely related to callers' hormonal quality (humans: Dabbs & Mallinger, 1999; Evans et al., 2008; giant pandas: Charlton et al, 2011) and mating success (humans: Apicella et al., 2009; fallow deer: Vannoni & McElligott, 2008).

Further evidence for the functional role of sexually dimorphic voice cues derives from perceptual investigations confirming that (natural and resynthesised) variation in sexually selected acoustic traits is salient to potential competitors and mates in many species. For example, acoustic signals of sexually mature males characterised by lower frequency components are typically seen as more attractive and dominant, giving males who produce them a competitive advantage (humans: Feinberg et al., 2008; red deer: Charlton, Reby & McComb, 2007; Reby et al., 2005). These studies complement work on faces and bodies in humans (Penton-Voak & Perrett, 2000; Thornill & Gangestad, 1999; Jackson, 1992; Mueller & Mazur, 2001; Mueller & Mazur, 1997), suggesting that acoustic and visual sexually selected traits signal common fitness-related dimensions and are used in conjunction by perceivers to assess the overall quality of the signaller (Candolin, 2003; Møller & Pomiankowski, 1993).

In addition to providing static indices to mate quality, sexually dimorphic traits can be dynamically controlled in order to gain a competitive advantage in agonistic and sexual interactions (Ohala, 1984). More specifically, because a larger animal is likely to win a physical confrontation over a smaller one, and attract mates, there is a strong selection pressure on animals to appear as large as possible (Ohala, 1984). Indeed, size exaggeration visual displays performed by aggressors, such as erecting hair or feathers, elevating the tail, arching the back or hunching the shoulders to appear larger, have been observed in several species of vertebrates (Davies and Halliday, 1978; Hauser, 1993; Ohala, 1984). In line with these observations, the "size code" hypothesis (Ohala, 1984) states that callers (including humans) may have evolved similar strategies in the acoustic domain, varying their sexually acoustic cues (in virtue of their relationship to body size) to exaggerate or downplay the impression of their size and related attributes, including dominance, aggression and competitive ability. The body-size projection principle has been confirmed in several species, whereby callers have been observed to dynamically adjust the formant spacing of their calls via dynamic elongation of their larynx to intimidate rivals and attract females (European red deer and wapiti: Fitch & Reby, 2001; fallow deer: McElligott et al., 2006; Mongolian gazelle: Frey et al., 2008). Since beginning work on my thesis, at least one study (Puts et al., 2006) has provided evidence for the use of a "size code" by human speakers, by showing that men lower their voice pitch when addressing a competitor that they perceive as less dominant than them.

Stemming from the "size code", the "gender code" proposed in this thesis complements evolutionary perspectives of voice gender differences by focusing on the fact that the same sexually selected voice cues of size also cue for typically human constructs such as gender, masculinity and femininity (Feinberg et al., 2008; Fraccaro et al., 2010; Hillenbrand et al., 2009). For example, because voice masculinity has been found to be typically associated with desirable mate qualities such as strength, social status, competence and trustworthiness (Sell et al., 2010; Vukovic et al., 2011), males may exaggerate their masculinity through the voice in order to accentuate these attributes, thus potentially gaining an advantage in acquiring mates or winning contexts.

However, it is also worth noting that, while in the literature the vocal expressions of size and gender are studied independently, they are largely interconnected. From an acoustic point of view, we have already observed that there is a natural overlap between size and gender: men, who speak with lower frequencies than women, are also on average bigger and more masculine than their female counterparts. However, physical relationships that exist between size and gender attributes, within as well as across genders, are yet to be explored (e.g. are larger men inherently more dominant and/or masculine than smaller men?). Indeed, my results on the physiological and acoustical bases of perceived masculinity (Study 3) suggest that body size and other biological markers of masculinity, such as salivary testosterone levels, are inter-related to some extent and expressed through the voice, though cued by individual voice features differentially (e.g. testosterone is not cued for by ΔF and height may be cued more by ΔF than F0). Future work is now needed to establish the extent to which the same voice features cue for size and related traits (e.g. dominance) as well as other markers of physical masculinity (e.g. facial hair and features). Moreover, in humans, size and gender attributes are also controlled through self-representation and behaviour, providing an added social dimension to the size–gender relationship: for example, it remains to be seen whether larger speakers may also consider themselves as more dominant and/or masculine, and whether they may engage in more dominant and/or masculine behaviours and activities than their smaller-bodied counterparts.

Even if size and gender features were to be independent dimensions, they may still be perceived by listeners in a more combinatory way. Indeed, size- and gender-associated meanings have relatively loose social definitions (Berger et al., 1972; Mac an Ghaill, 1996) and, as a result, people's conceptions of these two dimensions may also overlap considerably (e.g. rating speakers' voices on size and masculinity may be perceived as the same task by listeners). To date, only one study (Pisanski et al., 2012) has attempted to establish the perceptual interdependence of size and gender ratings, and found that listeners' perceptions of speakers' size overlap to a large extent, though do not perfectly coincide with their perceptions of masculinity and femininity (men and women with lower frequencies were rated as sounding larger and being more masculine). However, it remains to be explored whether these perceptions capture the real associations between speakers' size and gender attributes. Additional work is now needed to understand how size and gender attributes interlink at a personal level, to what degree they are both expressed in the voice and to what extent they are equated at a perceptual level.

Variation in sexually dimorphic voice cues: adding a social perspective

While acknowledging the important contribution of the evolutionary perspective in understanding human voice dimorphism, as emphasised by the size code theory (Ohala, 1984), I also argue that, by reducing sex differences, including voice differences, to ancestral mating strategies, evolutionary psychology fails to recognise the social nature of human sexuality. Acknowledging the fact that sexuality and gender in humans are to a large extent socially and culturally constructed dimensions enables a better understanding of the complexity and diversity of human gender voice differences. For example, different societies at different times define appropriate and desirable traits for men and women beyond actual evolved dispositions based on the notion of reproductive success (Feingold, 1994; Bem, 1983; Johannesen-Schmidt & Eagly, 2002). Accordingly, I hypothesise that individuals may vary the expression of gender through the voice, like other types of gendered-behaviour, as they routinely navigate across social expectations of girl/boy and woman/man, revising and maintaining their gender identities as necessary (Lindsey & Christy, 2011). Indeed, I deliberately chose the term "gender code" rather than “sex code” in framing my hypothesis as an explicit recognition that physiological as well as social differences may shape the acoustic diversity that characterises differences between male and female voices (and within one’s gender).

Several studies have evidenced that humans’ capacity to learn through observation and imitation of responses in others (Gergely, & Csibra, 2005) is central to the acquisition and maintenance of gender-typed knowledge and behaviour (Bussey & Bandura, 1999; Losin, Iacoboni, Martin & Dapretto, 2012). This means that individuals can exercise personal agency in expressing characteristics of what it is to be a “man” or a “woman” attributed by a given society and at a given time, although social and biological pressures may enhance or constrain agency (Archer, 1984; Conway et al., 1996). In line with this perspective and central to the “gender code”, is the observation that, among terrestrial mammals, humans have unique advanced vocal control and imitation abilities (Fitch, 2005). While it is generally accepted that the primary function of these abilities is to enable speech acquisition and production (Fitch, 2000, 2005), they also enable the sophisticated control of the quality of our voice in a variety of social contexts. For example, in the introduction to this thesis I pointed out that speakers have been found to vary their voice gender in acting contexts, or when complying with specific gender roles, as exemplified by male-to-female transsexual voices, features of “gay” speech, and language-specific between and within voice gender differences. I have also shown that, from early childhood, individuals are able to spontaneously vary the gender of their voice when asked to do so (Studies 4 and 5), as well as in response to cultural stereotypes (e.g. male actors projecting “effeminate” portrayals of homosexual men by feminising their voices – Study 8).

Gender-specific behaviours are acquired during early development as children learn to associate behaviours and appearance with one’s gender by observing others, and model their own behaviour accordingly. For example, children have been found to pay more attention to and remember toys if labelled for their gender (Bradbard et al., 1986), and to distort their memories when counter-stereotypical information on individuals was presented in order to fit them in the stereotype (Martin & Halverson, 1983; Liben & Signorella, 1993). Moreover, children have been found to look more at same-sex role models and remember more about them than about opposite-sex models (Slaby & Frey, 1975). While the acquisition of gender-specific voice behaviours remains to be studied, I have shown that children’s control of their voice gender reflects a conventionalised use of sex dimorphisms that are present in adult voices (lowering voice frequencies to sound more like a "boy" and raising them to sound more like a "girl" – Study 4). Given the absence of anatomical differences before puberty, these results suggest that individuals may learn from childhood how to "sound" like a man or a woman by observing and imitating adult same-sex models, as previously observed in other aspects of children's gender-typed behaviour (Biernat, 1991).

Finally, the social environment is central in shaping individuals' expression of their gender attributes. For example, it has been highlighted that adults view and treat boys and girls differently from infancy (Cassano, Zeman & Sanders, 2014; Fitch & Anderson, 2014), and the adoption of traditional gender roles is later on supplemented by teachers (Cahill & Adams, 1997; Hyde, Fennema, Ryan, Frost & Hopp, 1990; Thorne, 1993), same-sex peers (Banerjee & Lintern, 2000) and models in the surrounding environment (Turner & Gervai, 1995; Stoneman, Brody & MacKinnon, 1986). In turn, individuals learn from early childhood to respond to these social pressures by regulating their behaviour in response to different contexts: for example, boys have been found to accentuate their masculine traits in the presence of their peers (Banerjee & Lintern, 2000). The development of voice gender may also be subject to similar socialisation pressures. For example, the present observations (Studies 2 and 7) that both adult and child listeners attribute more masculine voices (lower ΔF) to more masculine children and more feminine voices (higher ΔF) to more feminine children, raise the question of whether cultural stereotypes and voice cues interact in shaping listeners' perceptions and behavioural responses to speakers. They also suggest that speakers may regulate their own behaviour to comply with, or elicit, gendered representations in others, as shown by male actors feminising their voices when playing homosexual characters (Study 8). Future work is needed to establish whether these gestures are also used in everyday life, for example whether individuals may vary the expression of their gender through the voice in line with their inner state, or situational context (e.g. type of audience, professional or personal interaction) and in the presence or absence of other cues to gender (e.g. body image, facial expressions, gait).

Potential impact

This research represents, to my knowledge, the first explicit and systematic investigation of vocal behaviour in relation to gender expression. It also offers, through the “gender code” hypothesis, a framework for organising existing findings and guides future research on describing and understanding voice gender variation, its origins, and its covariation with gender expression in general. As such, the outcomes of this research have the primary aim of advancing the field of human vocal communication, by improving our understanding of individuals’ ability to alter the gender-related characteristics of their voices in order to vary the expression of their gender, and the perceptual relevance of this behaviourally mediated variation to listeners.

By emphasising the role of pitch and resonance manipulations in the expression of voice gender, and relating those to underlying articulatory behaviours, the present results aim to scientifically contribute to the knowledge of professional voice coach and voice therapists, with potential implications for the design or enhancement of behavioural techniques used in the treatment of voice dysphoria or transitional voice change.

My research also sets the framework for future studies investigating the contribution of voice variation to gender expression: as dynamic changes to one’s voice can be objectively and quantifiably measured, future work could focus on the voice as an objective marker of gender identity development. This will be of significant interest to the wider scientific community (from core areas of developmental science through to interfaces with social and gender studies) as well as practitioners (speech therapy and coaching, education).

Finally, by showing that listeners and speakers apply the gender code to gender-stereotypic personality traits and roles, the present research highlights the need to investigate the expression of voice gender in different social contexts and with other modes of human communication and perception. The knowledge that will derive from this work might contribute to a more comprehensive understanding of how gender shapes human social interactions. More specifically, it could constitute an important step towards deconstructing simplistic and stereotyped representations of gender and sexual orientation, as well as helping individuals to understand (and possibly control) a key aspect of the social expression of their identity.

References

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter–gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1659), 1077–1082.

Archer, J. (1984). Gender roles as developmental pathways. *British Journal of Social Psychology*, *23*(3), 245–256.

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408.

Bee, M. A., Perrill, S. A., & Owen, P. C. (1999). Size assessment in simulated territorial encounters between male green frogs (Rana clamitans). *Behavioral Ecology and Sociobiology*, *45*(3-4), 177–184.

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs*, 598–616.

Berger, J., Cohen, B. P., & Zelditch Jr, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 241–255.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: a developmental analysis. *Journal of Personality and Social Psychology*, *61*(3), 351–365.

Bradbard, M. R., Martin, C. L., Endsley, R. C., & Halverson, C. F. (1986). Influence of sex stereotypes on children's exploration and memory: A competence versus performance distinction. *Developmental Psychology*, *22*(4), 481.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676.

Cahill, B., & Adams, E. (1997). An exploratory study of early childhood teachers' attitudes toward gender roles. *Sex Roles*, *36*(7-8), 517–529.

Candolin, U. (2003). The use of multiple cues in mate choice. *Biological Reviews*, *78*(4), 575–595.

Cassano, M. C., Zeman, J. L., & Sanders, W. M. (2014). Responses to Children's Sadness: Mothers' and Fathers' Unique Contributions and Perceptions. *Merrill-Palmer Quarterly*, *60*(1), 1–23.

Charlton, B. D., Keating, J. L., Kersey, D., Rengui, L., Huang, Y., & Swaisgood, R. R. (2011). Vocal cues to male androgen levels in giant pandas. *Biology Letters*, *7*(1), 71–74. doi:10.1098/rsbl.2010.0582

Charlton, B. D., Reby, D., & McComb, K. (2007). Female red deer prefer the roars of larger males. *Biology Letters*, *3*(4), 382–385.

Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2009). The information content of giant panda,< i> Ailuropoda melanoleuca</i>, bleats: acoustic cues to sex, age and size. *Animal Behaviour*, *78*(4), 893–898.

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences*, *27*(4), 801–804.

Darwin, C., (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

Davies, N. B., & Halliday, T. R. (1979). Competitive mate searching in male common toads,< i> Bufo bufo</i>. *Animal Behaviour*, *27*, 1253–1267.

Dixson, A. F. (2012). *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans*. Oxford University Press.

Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biological Psychology*, *72*(2), 160–163.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4), 783–788.

Fant G (1960) *Acoustic Theory of Speech Production.* The Hague: Mouton & Co.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior*, *29*(4), 233–241.

Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological Bulletin*, *116*(3), 429.

Fitch, W. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1477), 1669–1675.

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology*, *8*(3).

Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldaschev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (Gazella subgutturosa). *Journal of Anatomy*, *218*(5), 566–585.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222.

Fitch, W. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267.

Fitch, W. T. (2005). The evolution of language: a comparative review. *Biology and Philosophy*, *20*(2-3), 193–203.

Fitch, A., & Anderson, A. (2014). Parental Gender Stereotypes: Effects on Children's Toy Selection and Emotional Responses.

Flaherty, J., & Richman, J. (1989). Gender differences in the perception and utilization of social support: Theoretical perspectives and an empirical test. *Social Science & Medicine*, *28*(12), 1221–1228.

Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldaschev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (Gazella subgutturosa). *Journal of Anatomy*, *218*(5), 566–585.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology*, *6*(4), 219–225.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, *6*(3), 463–481.

Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulation in the production of rhesus monkey,< i> Macaca mulatta</i>, vocalizations. *Animal Behaviour*, *45*(3), 423–433.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2011). Voice correlates of mating success in men: examining 'contests' versus 'mate choice' modes of sexual selection. *Archives of Sexual Behavior*, *40*(3), 551–557.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender Comparisons of Mathematics Attitudes and Affect. *Psychology of Women Quarterly*, *14*(3), 299–324.

Jackson, L. A. (1992). *Physical appearance and gender: Sociobiological and sociocultural perspectives*. SUNY Press.

Johannesen–Schmidt, M. C., & Eagly, A. H. (2002). Another look at sex differences in preferred mate characteristics: The effects of endorsing the traditional female gender role. *Psychology of Women Quarterly*, *26*(4), 322–328.

Liben, L. S., & Signorella, M. L. (1993). Gender-schematic processing in children: The role of initial interpretations of stimuli. *Developmental Psychology*, *29*(1), 141.

Lindsey, L. L., & Christy, S. (2011). *Gender roles: A sociological perspective*. Pearson Prentice Hall.

Losin, E. A. R., Iacoboni, M., Martin, A., & Dapretto, M. (2012). Own-gender imitation activates the brain's reward circuitry. *Social Cognitive and Affective Neuroscience*, *7*(7), 804–810.

Mac an Ghaill, M. (1996). *Understanding masculinities: Social relations and cultural arenas*. McGraw-Hill International.

Martin, C. L., & Halverson Jr, C. F. (1983). The effects of sex-typing schemas on young children's memory. *Child Development*, 563–574.

McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (Dama dama) to lower their formant frequencies. *Journal of Zoology*, *270*(2), 340–345.

Møller, A. P., Soler, M., & Thornhill, R. (1995). Breast asymmetry, sexual selection, and human reproductive success. *Ethology and Sociobiology*, *16*(3), 207–219.

Muller, U., & Mazur, A. (1997). Facial dominance in Homo sapiens as honest signaling of male quality. *Behavioral Ecology*, *8*(5), 569–579.

Mueller, U., & Mazur, A. (2001). Evidence of unconstrained directional selection for male tallness. *Behavioral Ecology and Sociobiology*, *50*(4), 302–311.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. doi:10.1159/000261706

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1476), 1617–1623.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947. doi:10.1098/rspb.2004.2954

Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris). *Journal of Experimental Biology*, *202*(20), 2859–2867.

Ryan, M. J., & Brenowitz, E. A. (1985). The role of body size, phylogeny, and ambient noise in the evolution of bird song. *American Naturalist*, 87–100.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Rueden, C. von, … Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, rspb20100769. doi:10.1098/rspb.2010.0769

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 849–856.

Stoneman, Z., Brody, G. H., & MacKinnon, C. E. (1986). Same-sex and cross-sex siblings: Activity choices, roles, behavior, and gender stereotypes. *Sex Roles*, *15*(9-10), 495–511.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive*

Sciences, *3*(12), 452–460.

Turner, P. J., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. *Developmental Psychology*, *31*(5), 759.

Vannoni, E., & McElligott, A. G. (2008). Low Frequency Groans Indicate Larger and More Dominant Fallow Deer (Dama dama) Males. *PLoS ONE*, *3*(9), e3113.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research*, *54*(4), 995–1010.

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L. M., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, *102*(1), 37–48.